arXiv:2603.20070v1 [math.ST] 20 Mar 2026

Franz-Parisi 势的单调性等价于低阶 MMSE 下界

Konstantinos Tsirkas* Leda Wang* Ilias Zadik*2026 年3月23日

摘要

数十年来，两种截然不同的方法对于我们理解统计估计的计算复杂性发挥了重要作用。统计物理文献通过 Franz–Parisi（FP）势的局部稳定性和单调性，预测算法困难性 [FP95，FP97], 而数学上严谨的文献则通过受限算法类别的局限性来刻画困难性，其中最突出的是低次多项式估计器 [HS17]。对于许多推断模型，这两种视角给出的预测惊人地一致，因此，如何在二者之间建立精确的数学联系，成为一个长期悬而未决的问题。

近期研究 [BEAH+22，CMZZ25] 在检测问题中探讨了这一问题，表明对于广泛的模型类别，低次多项式的表现受一个涉及退火 FP 势的面积判据支配。尽管这为两个框架搭建了严谨的桥梁，但由此得到的判据不同于物理学文献传统上强调的基于单调性的条件，反映了检测与估计任务之间的根本区别。

在本文中，我们证明，对于信噪比为 λ 的一大类高斯加性模型（GAM），在估计问题中，低次多项式的能力等价于退火 FP 势的单调性。特别地，在 GAM 的低次猜想成立的前提下，我们的结果表明，这些模型的多项式时间极限可直接由退火 FP 势的单调性推出；这一结论在概念上与可追溯至 20 世纪 90 年代的物理学文献预测相符。具体而言，令 $\mathcal{F}_{\mathrm{ann}}\,\lambda$ 表示退火 FP 势，我们证明，最优的 D 次相关性对于所有 $D = O(\mathrm{poly}(\log n))$ 均满足如下近似不动点方程，

$$\mathrm{Corr}_{P_0}^{\leqslant D}\left(\lambda + \frac{d}{dq}F_{\mathrm{ann},\lambda}\Big|_{q=q(D)}\right)^2 \approx q(D),$$

其中 $q(D)$ 表示先验分布中两次独立抽样所得重叠的 e^{-D}分位数。特别地，条件$\frac{d}{dq}F_{\mathrm{ann}}\,\lambda\Big|_{q=q(D)} \geqslant 0$ 等价于所有次数不超过D 的多项式估计量，在信噪比 $q(D)$ 下所能达到的相关性至多为 λ。这建立了首个物理学启发的单调性判据与低次估计下界之间的严格等价性，也可视为经典 I–MMSE 关系 [GSV05]的低次类比。

*统计与数据科学系，耶鲁大学。电子邮箱：{kostas.tsirkas,leda.wang,ilias.zadik}@yale.edu

目录

1 引言 4
1.1 主要贡献 .5

2 主要结果：FP 单调性与低阶 MMSE 的等价性 8
2.1 递减 FP 得出的低阶相关性上界.8
2.2 递增 FP 得出的低阶相关性下界 .9
2.3 离散先验的等价性 .10
2.3.1 递减 FP 得出的低阶相关性上界（离散情形） 10
2.3.2 递增 FP 得出的低阶相关性下界（离散情形） 11
2.4 证明思路 .12

3 应用 13
3.1 张量主成分分析模型 .13
3.1.1 张量主成分分析（高斯先验） .13
3.1.2 稀疏张量主成分分析（拉德马赫稀疏先验） .13
3.2 高斯混合模型 .14
3.2.1 稀疏聚类 .14

4 结论 15

5 证明所需的其他符号说明 17

6 证明递增 FP 蕴含 MMSE 下界（定理 2.7 的证明） 17
6.1 累积量背景.17
6.1.1 关键引理 .18
6.2 上界的证明 .20
6.3 关键引理的证明 .22
6.3.1 引理 6.9 的证明 .22
6.3.2 命题 6.10 的证明 .23
6.3.3 引理 6.11 的证明 .24
6.3.4 引理 6.12 的证明 .26
6.4 辅助引理的证明 .27

7 递减 FP 蕴含 MMSE 上界（定理 2.9 的证明） 28
7.1 Hermite 背景 .29
7.2 关键引理 .30
7.3 下界的证明.31
7.4 关键引理的证明 .34

8 具有正累积量的两大类 GAM 36
8.1 两类 GAM.36
8.1.1 P_1 具有非负且超乘性矩.36
8.1.2 P_1 具有非负累积量.37
8.2 引理 8.2 的证明.38
8.3 引理 8.5 的证明.39

9 修正贝塞尔函数背景 40
9.1 第二类修正贝塞尔函数.41
9.2 高斯内积的密度 .42

10 应用证明 4210.1 稀疏张量主成分分析的证明.4310.1.1 定理 3.2 的证明 .4410.1.2 引理 10.2 的证明 .4510.1.3 引理 10.3 的证明 .4610.1.4 引理 10.4 的证明 .5110.1.5 引理 10.5 的证明 .5110.2 张量主成分分析的证明 .5310.2.1 定理 3.1 的证明 .5310.2.2 引理 10.13 的证明 .5410.2.3 引理 10.14 的证明 .5510.3 稀疏聚类的证明 .55

10.3.1 符号说明.5510.3.2 辅助引理 .5610.3.3 定理 3.3 的证明 .5610.3.4 引理 10.15 的证明 .5710.3.5 引理 10.16 的证明 .59

10.4 辅助引理 .6610.4.1 引理 10.25、10.26 的证明 .71A A 的失效

4A.1 算法阈值 .74A.2 退火 FP 势与低阶 MMSE 之间的等价性 . . . . . 75A.2.1 原始模型与截断模型之间低阶 MMSE 的等价性 . . . . . 75A.2.2 原始模型与截断模型之间退火 FP 势单调性的等价性 .79A.2.3 “容易”区域中的小低次数 MMSE .82A.2.4 综合起来 .84A.3 淬火 FP 势在“低次多项式容易”相中不是递减的 .85A.3.1 关键引理 .85A.3.2 定理 A.10 的证明 .88

B 辅助引理的证明 89

1 引言

数十年来，人们一直在努力理解贝叶斯统计估计任务的计算复杂性。由于这类任务具有平均情形的性质，要基于标准复杂性理论假设（例如 $\mathcal{P} \neq \mathcal{NP}$）刻画其困难性相，似乎远远超出了我们当前的数学能力。因此，来自不同背景的研究者提出了不同的方法来探讨这些问题。

针对低次多项式的最小均方误差下界 理论统计学和理论计算机科学中的一个重要研究方向，是探究受限类别估计量的局限性，这类研究通常被称为无条件下界。在这些框架中，低次多项式估计器在预测计算极限方面发挥了尤为核心的作用。尽管在估计问题中研究这一主题颇具挑战性， [SW22] 的开创性工作为如何界定某些高斯加性模型（GAM）中所有低次多项式的最小均方误差（MMSE）提供了一些关键技术。在这些模型中，对于某个先验分布 P_0 在 R^n 上，给定某个信噪比 λ,

$$Y = \sqrt{\lambda} X + Z, \tag{1.1}$$

对于信号 $X \sim P_0$和独立噪声 $Z \sim N(0, I_N)$。此后，更多研究为其他类别的高斯加性模型或伯努利模型给出了最小均方误差下界，例如 [SW25、LZ23、LG24、EGV25a]。值得强调的是，人们认为低次多项式在统计任务中具有强大能力，这一观点最终促成了一个颇具影响力的“低次猜想”，其针对的是类似的检测（假设检验）问题 [HS17], 该猜想认为，对于许多足够“良好”的检测问题，次数为 $D = O(\text{对数 } N)$的多项式检验，其表现可与最优多项式时间算法相匹敌。尽管目前尚未针对统计估计正式提出这一猜想的直接对应版本，但人们普遍预期，很快就会提出相应的猜想，至少适用于高斯加性模型。

Franz-Parisi 势的单调性 关于统计推断计算复杂性的另一种极具影响力的观点来自统计物理学界（例如，参见 [ZK16] 的综述）。从这一物理学视角来看，由于可以通过从贝叶斯后验分布中采样，得到实现最小均方误差（MMSE）的最优估计量，因此，最优的时间高效算法应对应于参数空间上一种符合物理直觉的可逆动力学，例如 Glauber 动力学或朗之万动力学，其平稳分布就是后验分布本身。因此，要理解这项任务的计算极限，归根结底就是分析这类动力学能否高效地从后验分布中采样，还是会陷入瓶颈（或“亚稳态”）。

在一项开创性工作中，Franz and Parisi [FP95，FP97] 引入了以重叠变量为参数的 Franz–Parisi（FP）势，用以预测这类动力学的行为。FP 势刻画了给定重叠的构型周围后验景观的局部几何，其形状可用于预测局部动力学会迅速混合，还是会陷入停滞。具体而言，物理学预测认为，当 FP 势不再递减时，动力学便会陷入亚稳态，而由此得到的估计量与信号的相关性，等于 FP 势达到第一个局部极小值时的重叠。

对于 GAM，FP 势具有以下形式（详情请参见 [BEAH+22] ），

$$\mathcal{F}_\lambda(q) := -\mathbb{E}_{X\sim\mu, Z\sim N(0,I_N), Y=\sqrt{\lambda}X+Z} \log \mathbb{E}_{X'\sim\mu} 1_{\langle X, X'\rangle = q} \exp\left(-\|Y - \sqrt{\lambda}X'\|_2^2/2\right), q \in [-1, 1]$$

由于 FP 势的形式复杂，物理学家通常会研究其易于处理的近似，以便做出具体预测。其中，两种近似尤为重要：复制对称（RS）近似和退火近似。研究表明，在某些尖峰矩阵模型中，RS 近似能够准确刻画近似消息传递（AMP）算法的行为：AMP 迭代的渐近相关性会收敛到 RS 势的第一个局部极小值的位置（例如，参见 [MW24，定理 1.1。] 及其中的参考文献）。退火近似通过对 FP 势应用詹森不等式并交换对数

与期望，得到一个显著简单得多的量，称为退火 FP 势（参见 [BEAH+22，命题 2。3]），其定义如下

$$\mathcal{F}_{\mathrm{ann},\lambda}(q) := -\log \mathbb{E}_{X,X'\sim\mu, X\perp\!\!\!\perp X'} 1_{\langle X,X'\rangle=q} \exp\left(\lambda\langle X, X'\rangle\right), q \in [-1,1].$$

一种数学关系？障碍与已有结果。令人惊讶的是，对于许多估计任务，低次多项式方法得出的计算困难性预测，与基于 Franz–Parisi（FP）势单调性的物理学观点所给出的预测一致。这种经验上的吻合自然引出了本研究所要探讨的核心问题，而该问题一直是这一领域尚未解开的谜题：

FP 势的单调性判据与低次多项式估计器所达到的最小均方误差之间，是否存在精确的数学联系？

遗憾的是，若试图利用式（1）中定义的原始（或淬火）FP 势来建立这种联系，结果会非常微妙，而且一般而言，若不施加实质性限制，这种做法并不成立。问题的根源在于，存在一些典型的估计问题——甚至属于高斯加性模型这一类——其淬火 FP 预测已被证明是不正确的。伯努利稀疏张量主成分分析就是一个显著例子，其中 $X = x^{\otimes t}$，对某个 $t \geqslant 2$，其中 x 从 $\{v \in \{0 \text{ 中均匀抽取，} 1/\sqrt{k}\}^N : \|v\|_0 = k\}$（另见第 A 节）[1]。在这种情形下，淬火 FP 势会在一个估计任务已知计算上容易的区域深处预测计算困难性，而低次多项式预测则被猜想是最优的（这是 [CSZ24 第 3 节所证明的瓶颈所带来的推论。3]）。不过，令人振奋的是，几年前 [MW24]，的一项启发性结果表明，对于 GAM 类中的（有偏、独立同分布）尖峰矩阵模型，近似消息传递（AMP）的最小均方误差表现与 $O(1)$ 次多项式估计器相同。由于 AMP 的渐近最小均方误差由 FP 势的复制对称（RS）近似刻画，这一结果令人期待，或许存在一种更广泛、更系统的理论，将基于 FP 的预测与低次估计极限联系起来。然而，尽管这一对应关系在概念上很有吸引力且结论精确，目前它仍仅适用于有偏、独立同分布的秩一尖峰矩阵高斯加性模型。

通过另一种 FP 判据建立检测情形下的等价关系 Bandeira 等人朝这一方向迈出了重要一步。 [BEAH+22] 他们证明，对于 GAM 的检测变体——其任务是在纯噪声与 GAM 之间进行区分——低次多项式检验的能力可由一种基于面积的判据刻画，该判据涉及式（1）[2]中定义的退火 FP 势。值得注意的是，使用退火 FP 势至关重要：对于稀疏张量主成分分析等问题，将相应的基于面积的判据应用于淬火 FP 势，仍无法与低次方法的预测相符。最近，[CMZZ25] 将这一联系大幅拓展到了 GAM 这一类模型之外。

尽管这些结果在检测情形下部分回答了我们最初的问题，但基于面积的判据是否与物理学文献强调的基于单调性的直觉相一致，仍未有定论。事实证明，退火 FP 曲线的单调性并非检测困难性的正确判据，这背后有着根本原因。许多 GAM——其中又包括稀疏张量主成分分析——都存在检测–估计间隙：检测可以在多项式时间内完成，而低次多项式方法进行估计则被猜想在计算上是困难的。在这种情况下，通过常规计算可以证明，单调性判据能够正确预测估计的困难性，却会错误地预测检测的困难性；相比之下，基于面积的判据能够正确刻画检测的低阶困难阈值。因此，如果要建立一种将 FP 势的单调性与低次多项式联系起来的相关理论，似乎就需要将讨论限定在估计情形中。

1.1 贡献

本工作的主要贡献在于，确立了低次多项式估计器所能达到的最小均方误差与退火 Franz–Parisi

[1]人们普遍认为，即使是带球面先验的简单张量主成分分析也是一个反例，但迄今为止，似乎只有淬火 FP 预测的低温或高温变体被严格证明失效 [AGJ20]。 [2]具体而言，该判据要考察退火 FP 势在原点附近的指数函数下的面积是否发散。

（FP）势的单调性之间精确的定量关系，该关系适用于一大类高斯加性模型（GAM）。我们将结果适用的这类 GAM 称为“低阶累积量非负”GAM。这一结果首次严格证明了两种先验上截然不同的计算困难性预测之间的等价性：一种源自统计物理中的退火 FP 势单调性判据，另一种则来自理论统计学和计算机科学中的低次 MMSE 下界。具体而言，我们的结果首次提供了一个形式化框架，在其中证明了物理学中的单调性判据能够刻画低次估计极限。

为陈述我们的结果，回顾一下：对于任意信噪比为 $\lambda > 0$、先验分布为 P_0、次数参数为 $D \in \mathbb{Z}_{>0}$ 的 GAM，次数为 D 的最小均方误差和次数为 D 的相关性定义为

$$\mathrm{MMSE}_{P_0}^{\leqslant D}(\lambda) := \inf_{f_1,\dots,f_n \in \mathbb{R}[Y]_{\leqslant D}} \mathbb{E}\left[\sum_{i=1}^{n} \big(f_i(Y) - X_i\big)^2\right], \mathrm{Corr}_{P_0}^{\leqslant D}(\lambda) := \sqrt{\sum_{i=1}^{n} \sup_{\substack{g_i \in \mathbb{R}[Y]_{\leqslant D} \\ \mathbb{E}[g_i(Y)^2]=1}} \mathbb{E}[g_i(Y)X_i]^2},$$

它们通过以下恒等式联系起来（见引理 6.8），$\text{最小均方误差}_{P_0}^{\leqslant D}(\lambda) = \mathbb{E}_{X\sim P_0}[\|X\|^2] - \big(\text{相关性}_{P_0}^{\leqslant D}(\lambda)\big)^2$.

我们的主要结果可非正式地概括如下。为简单起见，我们假设重叠 $\langle X, X'\rangle$（其中 X, X' 独立同分布于 P_0）服从连续分布，且其概率密度函数可微，因此退火 FP 势的导数有定义。

定理 1.1 (非正式表述，见定理2.7、2.9。)。 对于任意先验分布为P_0的低阶累积量非负 *GAM*，对于任意次数为 $0 < D \leqslant$ 多项式(对数 n) 以及任意信噪比 $\lambda > 0$，最优次数为D 的相关性满足

$$\big(\mathrm{Corr}_{P_0}^{\leqslant D}\big)^2 \left(\lambda + \frac{d}{dq}\mathcal{F}_{\mathrm{ann},\lambda}(q)\Big|_{q=q(D)}\right) \approx q(D),$$

其中，该近似仅隐去 n 中的多对数因子，而 $q(D)$ 表示重叠 e^{-D} 的 $|\langle X\text{-分位数}, X'\rangle|$（见定义 2.1）。

一个直接推论给出了物理学单调性判据与低阶估计困难性之间的精确等价性。

推论 1.2（非正式）。 对于任意具有先验分布 P_0 和信噪比 λ的“低阶累积量非负”*GAM*，以及任意 $0 < D \leqslant$ 多项式(对数 n)，以下命题在 n 的多对数因子意义下等价：

$$\frac{d}{dq}\mathcal{F}_{\mathrm{ann},\lambda}(q)\Big|_{q=q(D)} \geqslant 0 \quad \Longleftrightarrow \quad \big(\mathrm{Corr}_{P_0}^{\leqslant D}(\lambda)\big)^2 \leqslant q(D).$$

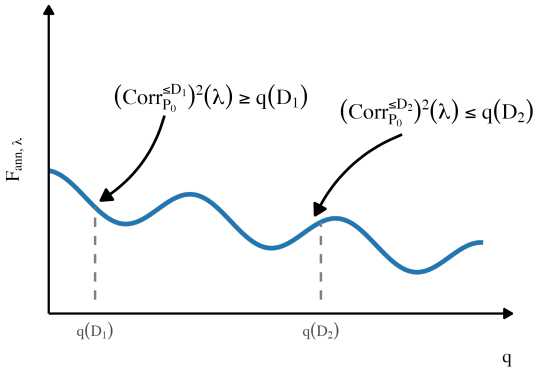
我们在概念层面简要描述这种对应关系。在基于 Franz–Parisi（FP）势的经典物理学图景中，预测是：由局部或可逆动力学建模的时间高效算法会被困在重叠值 q_1 处，该值对应于 FP 势的第一个局部最小值；而拥有无限计算资源的算法最终会到达重叠值 q_2处的全局最小值。

我们的结果证明，这一图景的精确量化类比在低次框架中是正确的（另见图1(a)）。假设我们用可使用的多项式估计量的最大次数 $D = 1、2、\dots\dots$来量化计算能力。对于任意信噪比 λ，我们考察退火 FP 势 $\mathcal{F}_{\mathrm{ann},\lambda}$ 在重叠取值 $q(D)$处的单调性。

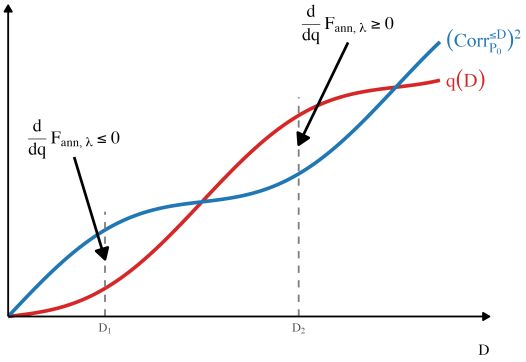
若$\frac{d}{dq}\mathcal{F}_{\mathrm{ann},\lambda}(q)\Big|_{q=q(D)} \leqslant 0$，那么物理学解释是，动力学下山，因此这在局部上是一个“容易”相。根据推论1.2，该相恰好对应于 $\big(\text{相关性}_{P_0}^{\leqslant D}(\lambda)\big)^2 \geqslant q(D)$，这意味着次数为$D$ 的多项式足够强大，能够达到至少为 $q(D)$的平方相关性。

反之，若 $\frac{d}{dq}\mathcal{F}_{\mathrm{ann},\lambda}(q)\Big|_{q=q(D)} \geqslant 0$，物理学预测表明，动力学必须爬山，因而会被困住，这标志着计算上的“困难”。与此完全对应地，根据推论1.2，该相对应于 $\big(\text{相关性}_{P_0}^{\leqslant D}(\lambda)\big)^2 \leqslant q(D)$，这表示次数为$D$ 的多项式无法达到平方相关性 $q(D)$。

从这个意义上说，重叠 $q(D)$ 处的物理学单调性预测，可以用低次估计的语言得到完整且严格的诠释。此外，这种对应关系是双向的：



(a) 这是退火势 $q \mapsto F_{\mathrm{ann}\,\lambda}(q)$ 随重叠参数 q 变化的示意图（信噪比 λ固定）。根据定理 1.1，选取曲线局部递减/递增处的点 $q = q(D_i), i = 1,$ 我们可以得出（相关性 $\overset{\leq D_i}{P_0})^2(\lambda)$（大致）大于/小于 $q(D_i)$



(b) 这是（相关性 $\overset{\leq D}{P_0})^2(\lambda)$ 和 $q(D)$ 随 D 变化的示意图（信噪比 λ 固定）。根据定理 1.1，我们可以利用这两条曲线的相对位置来确定 $\mathcal{F}_{\mathrm{ann}}$ 导数的符号。

图 1：我们等价关系的示意图。

低阶相关性函数随 D 变化的规律，可以反过来转化为退火 FP 势的精确单调性性质（见图 1(b)）。

注 1.3。 我们的结果有一个关键的技术特点，即 e^{-D} 分位数 $q(D)$ 在重叠$|\langle X, X' \rangle|$中所起的作用。如上所述，这些分位数提供了精确的参数化方式，将退火 FP 势递增/递减处的重叠值映射到相应的低阶 MMSE 上界/下界。我们强调，重叠分位数在退火 FP 框架中的重要性最早由 [BEAH+22] 指出；他们大致证明了，在区间 $[-q(D), q(D)]$ 上，$\exp(-\mathcal{F}_{\mathrm{ann},\lambda}(q))$ 曲线下方面积的有界性刻画了低阶检测困难性。

注 1.4。 虽然定理 1.1 针对连续先验分布给出，但我们也通过用适当的离散差分算子替代退火 FP 势的导数，证明了离散先验下的类似结果。

我们将关于 GAM 的精确技术假设，尤其是低阶累积量非负性的假设，留待论文正文中说明。不过这里指出，近期文献研究的若干典型 GAM 均满足这些假设。作为示例，我们将主定理应用于现有结果，并通过直接的黑箱式框架应用证明新的最先进低次 MMSE 下界。在所有情况下，证明都归结为对 $q(D)$ 的 $|\langle X, X' \rangle|$ 的重叠分位数进行界定。

特别是，我们的方法在低于猜测的算法阈值时，为以下问题得到了紧的 $O($对数 $n)$-次 MMSE 困难性结果：(a) 先验分布为 $X = x^{\otimes r}$ 的张量主成分分析，其中 x 的坐标服从独立同分布的高斯分布（第 3.1.1 节），或服从独立同分布的 $\mathrm{Rad}(k/n)$ 分布，且 $\tilde{\omega}(\sqrt{n}) = k = o(n)$（第 3.1.2 节）；以及 (b) 稀疏聚类模型，其中 $X = \xi\mu^\top$，ξ 的元素服从独立同分布的 $\mathrm{Rad}(1/2)$ 分布，而 μ 的坐标独立同分布，分布为标准高斯与 $\mathrm{Ber}(s/p)$ 随机变量的乘积，且 $\tilde{\omega}(\sqrt{p}) = s = o(p)$（第 3.2.1 节）。

最后，我们探讨一个自然的问题：我们的结果能否推广到淬火 FP 势？这一探讨很有依据，因为正如上文所述，在物理学中，退火 FP 势常被视为其淬火对应物的一个易于处理的替代方案（淬火 FP 势往往难以计算）。虽然这一推广很可能适用于我们研究涉及的许多 GAM，但必须强调的是，存在一些 GAM，在这些模型中我们的等价性成立，而淬火 FP 势的单调性却与低阶 MMSE 的行为不一致。具体来说，有些情况下，退火 FP 势的单调性能够成功刻画低阶

最小均方误差的行为，而淬火 FP 势的单调性却不能。我们在第 A 节介绍并分析了一个具体的反例（Rademacher稀疏三阶张量主成分分析的截断版本）。我们认为，退火 FP 势这种出人意料的“计算成功”——在预测算法困难性方面优于淬火势——是未来研究中一个重要的开放问题。

2 主要结果：FP 单调性与低阶最小均方误差之间的等价性

如引言中所述，我们的工作极为关注来自 GAM 先验分布 P_0 的两个独立同分布副本之间重叠的分位数。为陈述我们的结果，我们首先定义两个先验分布独立同分布抽样之间重叠绝对值的分位数函数 $\langle X,\ X'\rangle$。

定义 2.1（分位数函数）。设 $X,\ X'$ 为来自先验分布 P_0的独立同分布抽样。对于任意 $D>0$，我们将分位数 $q(D)$ 定义为 $q(D):=\inf\left\{y\in\mathbb{R}\mid \mathbb{P}(|\langle X,\ X'\rangle|\leqslant y)\geqslant 1-e^{-D}\right\}$.

为清晰起见，我们首先给出以下情形下的等价性结果：P_0 在 $\mathbb{R}^n$. 中具有连续可微概率密度函数。随后，我们将在最后一个小节，即第2.3节，说明在以下情形下几乎相同的结论（实际上，证明也几乎相同）：P_0 是离散分布。

2.1 由递减 FP 得出的低阶相关性上界

我们利用退火 FP 的单调性证明低阶相关性的一个上界。首先，我们准确描述对 GAM 所作的一系列假设。我们先介绍一个来自概率论的概念。

定义 2.2 (次 Weibull(ϕ))。 [例如，参见 [VGNA20]] 固定 $\phi>0$。称实值随机变量 X 为阶数为 ϕ 的次 Weibull 随机变量，$K>0$ 如果存在常数 $C_\phi\in(0,\ \infty)$ 使得$\left(\mathbb{E}|X|^p\right)^{1/p}\leqslant C_\phi K p^{1/\phi}$, 对所有 $p\geqslant 1$均成立。

注释 2.3（次 Weibull 随机变量的例子）。次 Weibull 随机变量涵盖许多熟悉的分布。例如：(i) 任意有界随机变量都是次 Weibull 随机变量(ϕ)，对每个 $\phi>0$ 均成立；(ii) 任意次高斯随机变量都是次 Weibull 随机变量(2)；(iii) 任意次指数随机变量都是次 Weibull 随机变量(1)。

我们研究先验分布满足低阶累积量非负性质的高斯加性模型（GAM）。该条件构成了我们的主要结构性假设，并辅以若干温和的正则性假设。

假设 2.4（低阶累积量非负 GAM）。固定一个普适常数 $C>0$。令 $(D_n)_{n\in\mathbb{N}}$为满足 $D_n\leqslant n^C$的一列正整数，并令 $(A_n)_{n\in\mathbb{N}}$ 为一列正实数。

先验分布为 P_0 的 GAM 满足我们的假设，其参数为 $(D_n)_{n\in\mathbb{N}}$、$(A_n)_{n\in\mathbb{N}}$和 $C>0$，并且满足以下条件。

(1)（D_n阶累积量非负）对于每个多重指标 $\alpha\in\mathbb{N}^n$，若 $|\alpha|\leqslant D_n$，则联合累积量满足 $\kappa_\alpha(X_1,\ldots,X_n)\geqslant 0$.

(2)（边缘矩的多项式增长）存在一个普适常数 $\phi>0$，使得对所有$i\in[n]$，边缘分布 X_i 都是次韦布尔(ϕ)。(3)（重叠分位数的受控衰减）存在一个普适常数 $C_1>0$，使得对所有 n，$q\big(D_n(\text{对数 }n)^2\big)\geqslant$最大值$\left\{(\text{对数 }n)^{-D_n/2},\ n^{-C_1}\right\}$。(4)（分位数稳定性）重叠分布满足（这里，$\mathbb{P}(\langle X,\ X'\rangle=t)$ 应理解为 $\langle X,X'\rangle$的概率密度函数），

$$\frac{d}{dt}\log\mathbb{P}(\langle X,X'\rangle=t)\Big|_{t=q(D_n)}\ \geqslant\ -\frac{A_n}{q(D_n(\log n)^2)\log n}. \tag{2.1}$$

假设 (1)–(4) 应理解为由一个主要结构条件——假设 (1)——以及三个辅助正则性条件组成；在 GAM 文献中考虑的大多数典范先验都满足这些辅助条件。

注释 2.5 （关于假设 (1)）。 假设 (1) 是支撑我们所述等价关系的主要技术条件，尤其是低次相关性上界。定理 2.7 （定理 2.9中的低次相关性下界）则不需要该条件。

许多自然先验分布都满足此条件。例如，张量先验 $X = x^{\otimes p}$ 满足假设 (1)：(a) 当 $D > 0$ 具有 x 独立同分布的高斯条目时，对所有 $D = O($；(b) 当 x 具有稀疏度为 k/n的独立同分布的伯努利–拉德马赫条目时，对数 $n)$ 。在第 8 节中，我们将展示一个范围广得多的先验类满足此条件，例如高斯稀疏聚类所用的先验 [EGV25a]。

假设 (2) 是标准且宽松的；参见注释 2.3。假设 (3) 也很弱：对于 D_n =ω(对数 n/对数对数 n)，它可化简为条件 $q(D_n) \geqslant n^{-C}$，其中 $C > 0$是某个普适常数；大多数感兴趣的先验分布都满足该条件。

注2.6（关于假设（4））。尽管假设（4）看起来可能有些技术性，但在大多数相关情形下，它是一个温和的条件。在许多 GAM 中，重叠 $\langle X,\ X'\rangle$ 是中心化的，由弱相关项组成——例如，在任意有限阶、坐标独立同分布的张量主成分分析中，或在稀疏张量主成分分析中。

在这类情形下，经过适当归一化后，重叠在分布上收敛到标准高斯分布的一个多项式：$\langle X,\ X'\rangle/a_n \xrightarrow{d} Z^p,\ Z \sim \mathcal{N}(0,\ 1)$，其中某个固定的 $p \in \mathbb{N}$。因此，对于任意 D_n =polylog(n)，分位数的表现如下（这也解释了该假设为何称为“分位数稳定性”）$q(D_n) = \Theta(a_n D_n^{p/2})$. 此外，通过类似的局部中心极限定理论证，我们自然也预期 $-\log \mathbb{P}(\langle X,\ X'\rangle = t)$ 接近 $(t/a_n)^{2/p}/2,$，二者共同表明

$$-\frac{d}{dt}\log \mathbb{P}(\langle X, X'\rangle = t)\Big|_{t=q(D_n)} = \Theta\left(\frac{D_n}{q(D_n)}\right),$$

这与假设（4）一致（其中允许有一些额外的松弛 A_n，应视为多对数因子，以及其他松弛多对数因子）。

现在我们可以陈述并证明第一个单向等价结果

定理 2.7 （递增的 FP 势意味着低次困难）。 设我们有一个先验分布为 P_0 且满足假设 2.4 的高斯加性模型，并且对每个 n， D_n， D_n' 正整数，满足 D_n' =$(1 + o(1))D_n$ 对数^2 n。

对于任意满足 $u + \frac{d}{dq}F_{\text{ann},\ u}\Big|_{q=q(D_n)} > 0$ 的 $u > 0$，我们有

$$\mathrm{Corr}^{\leqslant D_n}_{P_0}\left(\frac{1}{A_n}\cdot\left(u + \frac{\mathrm{d}}{\mathrm{d}q}\mathcal{F}_{\mathrm{ann},u}|_{q=q(D_n)}\right)\right)^2 \leqslant 2q(D_n'). \tag{2.2}$$

特别地，对于任意信噪比 $\lambda = \lambda_n > 0$，如果 n 足够大，我们有 $\frac{d}{dq}F_{\text{ann},\ \lambda}|_{q=q(D_n)} \geqslant 0$ 意味着相关性 $\mathrm{Corr}^{\leqslant D_n}_{P_0}\left(\frac{1}{A_n}\lambda\right)^2 \leqslant 2q(D_n')$.

Th定理的证明推迟到第6节；另请参见第2.4 节中的证明概要。

2.2 基于递增 FP 的低次相关性下界

现在我们转向等价性的另一面。这一次，只需对 GAM 作出温和的假设即可，尤其是这里不需要关于累积量非负性的假设。

在继续之前，我们先陈述本节将用到的假设。

假设 2.8。固定常数 $c, c_1, c_2, C, \kappa > 0$，一个正整数序列 $(D_n)_{n\in\mathbb{N}}$，使得 $D_n = \omega(1)$，以及一个正数序列 $(B_n)_{n\in\mathbb{N}}$，使得 $B_n = \omega(\max\{n^{-C}, e^{-CD_n}\})$。若满足以下条件，则具有先验分布 P_0 的 GAM 对参数 $(D_n)_{n\in\mathbb{N}}, (B_n)_{n\in\mathbb{N}}$ 和 $c, c_1, c_2, C, \kappa > 0$ 满足该假设。

(1)（固定分位数阶）对于所有 $D > 0$, $c_1 B_n D^\kappa \leqslant q(D) \leqslant c_2 B_n D^\kappa$。

(2) （分位数 重叠的概率密度函数满足， $\frac{d}{dt} \log \mathrm{P}(\langle X,\ X' \rangle = t)\big|_{t=q(D_n)} \leqslant -\frac{c}{q(D_n}$ ）.

假设（1）、（2）相当宽松，文献中的大多数 GAM 都满足这些假设。事实上，这些假设背后的动机完全源于假设（4）中所述的内容，详见备注 2.6 关于假设 2.4 。有趣的是，请注意，本节需要采用与假设（4）相反的不等式方向。

现在我们可以陈述这一方向的主要结果了。

定理 2.9 (递减 FP 势意味着低次问题容易)。 对于满足假设 2.8、具有序列 $(B_n)_{n\in\mathbb{N}}$、 $(D_n)_{n\in\mathbb{N}}$ 及常数 c、c_1、 c_2、 C、 $\kappa > 0$ 的任意 *GAM*，存在一个常数 $C' > 0$，使得以下结论成立。对于任意满足 $u + \frac{d}{dq} F_{\text{ann},}$ $u\Big|_{q=q(D_n)} > 0$ 的 $u > 0$，我们有

$$(\mathrm{Corr}_{P_0}^{\leqslant D_n})^2 \left(u + \frac{d}{dq} \mathcal{F}_{\text{ann},u}\Big|_{q=q(D_n)} \right) \geqslant \frac{C'}{D_n^{2\kappa}} q(D_n).$$

特别地，对于任意信噪比 $\lambda > 0$， $(\text{相关性}_{P_0}^{\leqslant D_n})^2(\lambda) \leqslant \frac{C'}{D_n^{2\kappa}} q(D_n)$ 蕴含 $\frac{d}{dq} F'_{\text{ann},\ \lambda}\Big|_{q=q(D_n)} \geqslant 0.$

Th该定理的证明推迟至附录A 7；证明概要亦见第 2.4 节。

2.3 离散先验的等价性

这里我们陈述离散先验分布 P_0的等价性结果。唯一的区别在于，对数$(\Pr(\langle X,\ X' \rangle = q))$ 的导数不再是良定义的，其中现在的 $\Pr(\langle X,\ X' \rangle = q)$ 是一个概率质量函数（PMF），因此我们改为考虑离散导数/差分算子。重要的是，除这一差别外，本节的所有陈述、假设和证明均与第 2 节中的相同。更具体地说，证明中唯一的区别是，为了应用这些定理，需要根据这个现在的离散导数算子来检验假设。

为方便起见，我们用 $f_{\text{ov}}(q) = \Pr(\langle X, X' \rangle = q)$ 表示从先验分布中独立抽样两次所得重叠的概率质量函数。定义离散导数算子时需要谨慎，因为 $f_{\text{ov}}(q)$ 的支撑集仅包含 q 的离散值。令 $\mathcal{S}_{\text{ov}} = \{q \in \mathrm{R} : f_{\text{ov}}(q) > 0\}$。现在，取任意函数序列$a_n : \mathcal{S}_{\text{ov}} \to \mathrm{R}_{>0}, n \in \mathrm{N}$，使得对所有 $q \in \mathcal{S}_{\text{ov}}$, $q + a_n(q) \in \mathcal{S}_{\text{ov}}$ 均成立。给定这一选择 $a_n, n \in \mathrm{N}$，我们定义作用于函数 $f : \mathcal{S}_{\text{ov}} \to \mathrm{R}$ 的离散导数算子 $\mathbf{\Delta}_{a_n}$，使得

$$\mathbf{\Delta}_{a_n}(f)(q) = \frac{f(q + \alpha_n(q)) - f(q)}{a_n(q)}, q \in \mathcal{S}_{\text{ov}}.$$

2.3.1 由递减 FP 得到的低阶相关性上界（离散情形）

与连续情形类似，要证明这一结果，我们需要一些假设；这些假设与假设 2.4相同，只需将导数替换为离散导数算子。

假设 2.10（低阶累积量非负 GAM（离散情形））。固定一个普适常数 $C > 0$。设$(D_n)_{n\in\mathbb{N}}$ 为满足 $D_n \leqslant n^C$ 的正整数序列，设 $(A_n)_{n\in\mathbb{N}}$ 为正实数序列，并取任意“速度”函数序列 $a_n : \mathcal{S}_{\text{ov}} \to \mathrm{R}_{>0}, n \in \mathrm{N}$，使得对所有 $q \in \mathcal{S}_{\text{ov}}$, $q + a_n(q) \in \mathcal{S}_{\text{ov}}$ 均成立。

假设先验分布为 P_0 的 GAM 满足以下条件，其参数为 $(D_n)_{n\in\mathbb{N}}$、 $(A_n)_{n\in\mathbb{N}}$、 $(a_n)_{n\in\mathbb{N}}$ 和 $C > 0$。

(1)（D_n阶累积量非负）对于每个满足 $|\alpha| \leqslant D_n$ 的多重指标 $\alpha \in \mathrm{N}^n$，其联合累积量满足

$$\kappa_\alpha(X_1, \ldots, X_n) \geqslant 0. \tag{2.3}$$

(2) (边缘矩的多项式增长) 存在一个普适常数 $\phi > 0$，使得对于所有$i \in [n]$，边缘分布 X_i 均为次韦布尔分布 (ϕ)。

(3) (重叠分位数的受控衰减) 存在一个普适常数 $C_1 > 0$，使得对于所有 n，$q\big(D_n(\text{对数 } n)^2\big) \geqslant$最大值 $\big\{(\text{对数 } n)^{-D_n/2},\, n^{-C_1}\big\}$.

(4) (分位数稳定性) 重叠分布满足（这里的 $\mathbb{P}(\langle X, X'\rangle = t)$ 应理解为 $\langle X, X'\rangle$ 的概率质量函数），

$$\boldsymbol{\Delta}_{a_n} \log \mathbb{P}(\langle X, X'\rangle = t)\big|_{t=q(D_n)} \;\geqslant\; -\frac{A_n}{q(D_n(\log n)^2)\log n}. \tag{2.4}$$

在此假设下，我们得到离散情形中的以下结果。

定理 2.11 （递增的 FP 势意味着低次困难（离散情形））。 设我们有一个先验分布为 P_0 且满足假设 *2.10* 的高斯加性模型，并对每个 n， D_n， D'_n 取满足以下条件的正整数$D'_n = (1+o(1))D_n$对数^2 n。

对于任意满足 $u > 0\, F_{\text{ann},\, u}\Big|_{q=q(D_n)} > 0$ 的 $u + \boldsymbol{\Delta}_{a_n}$，我们有

$$\operatorname{Corr}^{\leqslant D_n}_{P_0}\left(\frac{1}{A_n}\cdot\big(u + \boldsymbol{\Delta}_{a_n}\mathcal{F}_{\text{ann},u}|_{q=q(D_n)}\big)\right)^2 \leqslant 2q(D'_n). \tag{2.5}$$

特别地，对于任意信噪比 $\lambda = \lambda_n > 0$，如果 n 足够大，我们有 $\boldsymbol{\Delta}_{a_n} F_{\text{ann},\, \lambda}|_{q=q(D_n)} \geqslant 0$ 蕴含 相关性 $\overset{\leqslant D_n}{P_0}\left(\frac{1}{A_n}\lambda\right)^2 \leqslant 2q(D'_n)$.

证明与定理 2.7 的证明相同，只需交换导数算子和离散导数算子，故略去。

2.3.2 由递增 FP 得到的低次数相关性下界（离散情形）

现在，将正向结果的假设修改如下。

假设 2.12。固定常数 $c, c_1, c_2, C, \kappa > 0$，一个正整数序列 $(D_n)_{n\in\mathbb{N}}$，使得 $D_n = \omega(1)$；一个正数序列 $(B_n)_{n\in\mathbb{N}}$，使得 $B_n = \omega(\max\{n^{-C}, e^{-CD_n}\})$；并任取一个“速度”函数序列 $a_n : \mathrm{S}_{\text{ov}} \to \mathrm{R}_{>0}, n \in \mathrm{N}$，使得对所有 $q \in \mathrm{S}_{\text{ov}}$, $q + a_n(q) \in \mathrm{S}_{\text{ov}}$。假定先验分布为$P_0$ 的 GAM 在参数 $(D_n)_{n\in\mathbb{N}}, (B_n)_{n\in\mathbb{N}}, (a_n)_{n\in\mathbb{N}}$ 和 $c, c_1, c_2, C, \kappa > 0$ 满足以下条件时成立。

(1)（固定分位数阶）对于所有 $D > 0$， $c_1 B_n D^{\kappa} \leqslant q(D) \leqslant c_2 B_n D^{\kappa}$。

（2）（分位数稳定性）重叠的概率质量函数满足，

$$\boldsymbol{\Delta}_{a_n} \log \mathbb{P}(\langle X, X'\rangle = t)\big|_{t=q(D_n)} \leqslant -\frac{c}{q(D_n)}.$$

现在我们已经准备好陈述这一方向在离散情形下的主要结果。

定理 2.13 （递减 FP 势意味着低次易解（离散情形））。 对于满足假设 *2.12* 的任意 *GAM*，以及序列 $(B_n)_{n\in\mathbb{N}}$、 $(D_n)_{n\in\mathbb{N}}$、 $(a_n)_{n\in\mathbb{N}}$和常数 c、 c_1、 c_2、 C、 $\kappa > 0$，存在一个常数 $C' > 0$，使得以下结论成立。对于任意满足 $u + \boldsymbol{\Delta}_{a_n} F_{\text{ann},\, u}\Big|_{q=q(D_n)} > 0$ 的 $u > 0$，我们有

$$(\operatorname{Corr}^{\leqslant D_n}_{P_0})^2\left(u + \boldsymbol{\Delta}_{a_n}\mathcal{F}_{\text{ann},u}\Big|_{q=q(D_n)}\right) \geqslant \frac{C'}{D_n^{2\kappa}} q(D_n).$$

特别地，对于任意信噪比 $\lambda > 0$， $(\text{相关性}^{\leqslant D_n}_{P_0})^2(\lambda) \leqslant \frac{C'}{D_n^{2\kappa}} q(D_n)$ 意味着 $\boldsymbol{\Delta}_{a_n} F'_{\text{ann},\, \lambda}\Big|_{q=q(D_n)} \geqslant 0.$

与上述证明类似，该证明只需交换导数算子与离散导数算子，即与定理 2.9 的证明完全相同，故此处省略。

2.4 证明思路

虽然等价定理的证明将留待后文给出，但我们先简要概述其核心思路。为简单起见，假设重叠分布具有连续且可微的密度。

一个关键的出发点是下面这个初等恒等式，它显著阐明了退火 FP 势在 GAM 中的作用：对于任意信噪比 $\lambda > 0$ 以及所有 $q \in [-1,\ 1]$，都有 $\lambda q + \mathcal{F}_{\mathrm{ann},\ \lambda}(q) = -\log \mathrm{P}(\langle X,\ X' \rangle = q)$.。特别地，在 $q = q(D)$ 处求导可得

$$\lambda + \frac{d}{dq}\mathcal{F}_{\mathrm{ann},\lambda}(q)\Big|_{q=q(D)} = -\frac{d}{dq}\log \mathbb{P}(\langle X, X' \rangle = q)\Big|_{q=q(D)}. \tag{2.6}$$

在假设 2.4 以及假设 2.8成立且次数 D =polylog(n)时，可以证明（2.6）右侧满足 $-\frac{d}{dq}$对数 $\mathrm{P}(\langle X,\ X' \rangle = q)\Big|_{q=q(D)} \approx \frac{1}{q(D)}$.结合前两步可知，要证明我们的主定理，实际上有趣地归结为：对于所有相关的 GAM 以及次数 $D = \mathrm{polylog}(n)$，$\left(\text{相关性}_{P_0}^{\leqslant D}\right)^2\left(\frac{1}{q(D)}\right) \approx q(D)$，也就是说，低阶相关性在 $q(D)$的尺度上满足近似不动点关系。

我们分别证明这两个不等式。对于上界 $\left(\text{相关性}_{P_0}^{\leqslant D}\right)^2\left(\frac{1}{q(D)}\right) \leqslant q(D)$，我们借助 [SW22], 的一项最新成果，该成果通常被称为詹森技巧，它可以用先验分布的累积量来给出低阶相关性的上界 P_0。这样一来，问题便归结为验证一个将 P_0 的累积量与重叠的分位数联系起来的不等式 $\langle X,\ X' \rangle$。证明利用了这样一个事实：重叠分位数可以通过对数矩母函数来控制，而对数矩母函数又可以通过累积量的泰勒展开来表示。证明细节颇为精妙，完整证明见第6节。

为了证明相应的下界，$\left(\text{相关性}_{P_0}^{\leqslant D}\right)^2\left(\frac{1}{q(D)}\right) \geqslant q(D)$，我们显式构造了一个次数为$D$ 的估计量，在对GAM施加温和条件的情况下，其平方相关性至少为 $q(D)$ 。一种自然的方法是分析最优的次数为D 的多项式；它对应于在 Y的边缘测度下，将后验均值投影到次数为D 的多项式所构成的线性空间上。

由于 Y不服从乘积测度，因此不存在已知的正交归一多项式规范基，文献通常认为分析这一投影“在技术上难以处理”——即使对于简单的 GAM 也是如此。这也是研究人员经常通过其他方式分析低阶 MMSE 的原因。在本工作中，我们构造了一个与后验均值的最优低阶投影非常接近的代理，它易于定义和分析，适用于我们在本工作中考虑的所有 *GAM*，而且本身也可能具有独立的研究价值。通过这一分析，我们得到了所需的低阶相关性紧下界。

该构造基于一种重要性采样方案：我们从先验分布 P_0 中抽取 M个独立样本 $X_1, \ldots, X_M$，并考虑如下形式的多项式估计量

$$p(Y) = \sum_{i=1}^{M} W(Y \mid X_i) X_i, \tag{2.7}$$

其中，权重 $W(Y \mid X_i) = W_D(Y \mid X_i)$ 是经过精心选择的D 次多项式，旨在最大化其与信号的相关性。选择这些权重的动机是让 $p(Y)$ 近似后验均值的低阶投影。关键在于，将厄米特多项式视为在 Y 的测度下构成正交基，并据此将其投影到这些多项式上。这导出了权重的规范选择 $W = W_D$ （见（7.12））， 并使得可处理的分析成为可能，从而建立所需的下界，完成对近似不动点关系的证明。值得注意的是，在某些模型中，将厄米特/傅里叶–沃尔什基视为在“植入” Y 下的近似标准正交基这一思路，近期已出现在低阶 MMSE 文献中 [CGGV25] （在 [CGGV25] 中，关注的并非 GAM，而是一些随机图模型）。即使在此情形下，这一构造的技术细节也颇为微妙，相关内容见第7节。

3 应用

为展示我们等价性结果的适用性，我们证明这些结果能够通过以下方式，得到近期文献中的一些已有及最新 MMSE 下界：（1）直接计算先验分布下重叠的分位数；（2）证明退火 FP 势的导数非负（因而属于“物理困难”问题）；以及（3）应用定理 2.7 （或离散情形下的定理 2.11 ），得出其低阶 MMSE 难以求解的结论。

我们强调，在这些证明中，技术性工作在于检验假设 2.4 和定理 2.7。其中，除假设（4）有时更具挑战性外，其余部分都相对容易检验。为验证假设（4），我们证明了若干关于重叠度密度的局部中心极限定理，这可视为一项可能具有独立研究价值的技术性贡献。所有证明均推迟至第 10节。

3.1 TensorPCA 模型

在此模型中，先验分布 P_0 具有如下形式： $X = \mathrm{vec}(v^{\otimes r})$ ，其中 $r > 0$为某个常数，$v \in \mathbb{R}^N$ 上的某个分布，以及 $n = N^r$。特别地，注意到在这种先验分布下，对于任意 $X = \mathrm{vec}(v^{\otimes r})$、 $X' = \mathrm{vec}((v')^{\otimes r})$，都有 $\langle X,\ X'\rangle = \langle \mathrm{vec}(v^{\otimes r}),\ \mathrm{vec}(v'^{\otimes r})\rangle = \langle v,\ v'\rangle^r$.

3.1.1 张量主成分分析（高斯分布先验分布）

在此模型中， $v \in \mathbb{R}^n$ 具有独立同分布的$\mathcal{N}(0,\ 1)$ 条目。此模型的统计阈值为 $\lambda_{\mathrm{IT}} = \widetilde{\Theta}(n^{1-r})$[MR14] ，而算法阈值的猜想为 $\lambda_{\mathrm{ALG}} = \widetilde{\Theta}(n^{-r/2})$ （例如参见 [HKP+17] 及其中的参考文献）。

事实证明，对于任意 $r \geqslant 1$，由于先验分布对任意2.4 都满足假设 $D_n = \tilde{o}(\sqrt{n})$、威布尔常数 $\phi = 2$ 和 $A_n = \Theta(D_n^{3/2}$对数$^{r+1}\, n)$，定理2.7适用于张量 PCA 模型。特别地，定理 2.7 推出以下结果。

定理 3.1。固定 $r \geqslant 2$。对于秩一张量尖峰模型中的高斯先验，我们证明，对所有$\lambda > 0$ 以及任意整数序列 $D_n = o(\sqrt{n}/$对数$^2\, n)$:

$$\mathrm{MMSE}_X^{\leqslant D_n}\left(\frac{1}{D_n^{3/2}\log^{r+1} n}\big(\lambda + \frac{\mathrm{d}}{\mathrm{d}q}\mathcal{F}_{\mathrm{ann},\lambda)}\Big|_{q=q(D_n)}\big)\right) = n^r - \tilde{O}\big((D_n n)^{r/2}\big) = (1+o(1))\mathrm{MMSE}_X^{\mathrm{trivial}}.$$

特别地，对于某个 $\lambda = \widetilde{\Theta}(n^{-r/2}) = \widetilde{\Theta}(\lambda_{\mathrm{ALG}}),$ ，我们有 $\mathrm{d}\ \mathrm{d}q\mathcal{F}_{\mathrm{ann},\ \lambda}\Big|_{q(D_n)} \geqslant 0$，因此

$$\mathrm{MMSE}_X^{\leqslant D_n}\left(\widetilde{\Theta}\Big(\frac{\lambda_{\mathrm{ALG}}}{D_n^{r/2}}\Big)\right) \geqslant (1+o(1))\mathrm{MMSE}_X^{\mathrm{trivial}}.$$

定理的证明推迟到第 10.2 节。

对于张量 PCA 设置（具有独立同分布的“稠密”先验），低阶 MMSE 下界也已在 [KMW24] 中针对拉德马赫先验得到，所用方法来自自由概率论，但仅适用于r 奇数情形。我们的方法能够针对高斯先验证明所有 r 情况下的低阶 MMSE 下界。

3.1.2 稀疏张量主成分分析（拉德马赫稀疏先验）

现在我们转向另一种张量主成分分析设置，它采用一种研究得十分透彻的离散先验。令 $k \in \mathbb{N}$ 满足 $k = n^{\beta+o(1)}$ 对某个 $\beta \in (1/2, 1)$成立。于是，稀疏（拉德马赫）张量 PCA 模型是指先验分布为 $v \in \mathbb{R}^n$ 且其元素独立同分布、服从 $\mathrm{Rad}(k/n),$的张量 PCA 模型，这意味着对所有 $i = 1,\ \ldots\ldots,\ N,\ \ v_i = 1$ 以概率 $k/(2n),\ v_i = -1$以概率 $k/(2n)$ 且 $v_i = 0$，否则。该模型的统计阈值为 $\lambda_{\mathrm{IT}} = \widetilde{\Theta}(k^{1-r})$ [LZ23] ，而该模型的算法阈值普遍猜测为$\lambda_{\mathrm{ALG}} = \widetilde{\Theta}(\min\{1,\ n^{r/2}/k^r\})$ [LZ23]。

事实证明，对于任意 2.7 定理2.11 （附录中的定理），其离散版本适用于稀疏张量主成分分析模型，因为稀疏拉德马赫先验分布满足假设 $r \geqslant 1,\ \beta \in (0,\ 1/2) \cup (1/2, 1),$其离散版本适用于 2.4 （具体而言，假设 2.10），当 $D_n = \lfloor (1-\beta)$对数 $n \rfloor$，且 $A_n =$

$\Theta(D_n^{2-r}$对数$^{(3r+2)/2}\, n)$。特别地，定理 2.7 的离散版本推出以下结果（下文的$\mathbf{\Delta}_{a_n}$ 对应于离散导数算子，详情请参见第 2.3 节）。

定理 3.2。 固定 $r \geqslant 2$。对于稀疏张量主成分分析模型，当所有信噪比 $\lambda > 0$ 和 $D_n = \lfloor (1-\beta)$对数 $n \rfloor$满足条件时，我们有以下结论。

If $\beta \in (\frac{1}{2}, 1)$, *then*

$$\mathrm{MMSE}^{\leqslant D_n}\left(\frac{1}{D_n^{r-2}\log^{(3r+2)/2} n}\cdot\left(\lambda + \mathbf{\Delta}_{a_n}\mathcal{F}_{\mathrm{ann},\lambda}\Big|_{q=q(D_n)}\right)\right) = k^r - \widetilde{O}\left(\frac{k^r}{n^{r/2}}\right) = (1+o(1))\mathrm{MMSE}^{\mathrm{trivial}}.$$

特别地，由于对于某个 $\lambda = \widetilde{\Theta}(n^{r/2}/k^r) = \widetilde{\Theta}(\lambda_{\mathrm{ALG}})$, $\mathbf{\Delta}_{a_n}F_{\mathrm{ann},\lambda}\Big|_{q(D_n)} \geqslant 0$，我们得出结论

$$\mathrm{MMSE}^{\leqslant D_n}\left(\widetilde{\Theta}(\lambda_{\mathrm{ALG}})\right) \geqslant (1+o(1))\mathrm{MMSE}^{\mathrm{trivial}}.$$

如果 $\beta \in (0, \frac{1}{2})$，那么

$$\mathrm{MMSE}^{\leqslant D_n}\left(\frac{1}{\log^3 n}\cdot\left(\lambda + \mathbf{\Delta}_{a_n}\mathcal{F}_{\mathrm{ann},\lambda}\Big|_{q=q(D_n)}\right)\right) = k^r - \widetilde{O}(1) = (1+o(1))\mathrm{MMSE}^{\mathrm{trivial}}.$$

特别地，由于对于某个 $\lambda = \widetilde{\Theta}(1) = \widetilde{\Theta}(\lambda_{\mathrm{ALG}})$, $\mathbf{\Delta}_{a_n}F_{\mathrm{ann},\lambda}\Big|_{q(D_n)} \geqslant 0$，我们得出结论

$$\mathrm{MMSE}^{\leqslant D_n}\left(\widetilde{\Theta}(\lambda_{\mathrm{ALG}})\right) \geqslant (1+o(1))\mathrm{MMSE}^{\mathrm{trivial}}.$$

定理的证明推迟至第10.1.1节。

据我们所知，这是针对该设定的首个紧致低阶 MMSE 下界。对于此模型，已有一些针对支撑恢复问题的估计下界 [LZ23]，该问题关注的是恢复 v的支撑集；但（a）这些结果似乎不适用于恢复带符号向量 x；（b）它们依赖于基于归约的困难性论证，而非直接的低阶 MMSE 下界。

3.2 高斯混合模型

在具有两个中心的高斯混合模型中，我们观测到 n 个位于 $\mathbf{R}^p$ 中的数据点，其中每个数据点以相同概率取自$\mathcal{N}(\mu_1, \sigma^2 I_p)$或$\mathcal{N}(\mu_2, \sigma^2 I_p)$其中 $\mu_1, \mu_2 \in \mathbf{R}^p$ 是这两个“中心”。我们将这 n 个观测作为矩阵 $Y \in \mathbf{R}^{n\times p}$ 的行收集起来。该模型的规范参数Δ^2 与两个中心之间的最小间距成正比，即 $\Delta^2 := \frac{\|\mu_1-\mu_2\|_2^2}{2\sigma^2}$.

我们关注平衡情形，其中两个中心为 μ 和 $-\mu$，对某个 $\mu \in \mathbf{R}^p$，且不失一般性，假设 $\sigma = 1$。等价地，令潜在标签 $\xi_i \in \{+1, -1\}$独立同分布，满足$\mathbf{P}(\xi_i = +1) = \mathbf{P}(\xi_i = -1) = 1/2,$，并假设 $Y_i = \xi_i\mu + z_i$，对 $z_i \overset{i.i.d.}{\sim} \mathcal{N}(0, I_p)$。在这种情况下，这恰好对应于 $n \times p$ 维的 GAM 设置，其中 $Y = X + Z$设置适用于 $\xi = (\xi_1, \ldots, \xi_n)^\top$，且 $X = \mathrm{vec}(\xi\mu^\top)$ 是行向量为 $Y_i \in \mathbf{R}^p$, $i = 1, \ldots, n$ 的 $n \times p$矩阵向量化后的结果。此外，在这种情况下 $\Delta^2 = 2\|\mu\|_2^2$.

3.2.1 稀疏聚类

在稀疏聚类中，假设均值 μ 是稀疏的。具体而言，存在一个未知集合 $J \subseteq [p]$，满足 $|J| \leqslant s$，使得 $\mu_j = 0$ 对所有 $j \notin J$均成立。我们将重点讨论 $s = p^{\alpha+o(1)}$这一情形，其中 $\alpha \in (1/2, 1)$. 在我们感兴趣的参数范围内，推测该模型存在一个算法阈值；按照我们的标度，该阈值为$\Delta^2 = \Delta^2_{\mathrm{ALG}} = \Theta(p/n)$ [LWB21]。

现在我们证明，在预测的算法阈值附近存在低阶 MMSE 下界。考虑定义在我们的中心上的以下先验分布 μ $-\mu$ [25] $\Delta > 0$ $b_j \sim$ $(s/p), j =$，以及，借鉴自 EGVb。令 和 为 $1, \ldots, p$ $g_j \sim N(0\ \frac{\Delta}{s})$ $j = 1$ $\ldots$ n $\mu_j = b_j g_j$ $j = 1$ $\ldots$ p. 独立同分布的 Ber，且 为独立同分布的 ，，，，据此我们设定：，，，注意，对于均值为$\|\mu\|_2^2$的这一选择， $\Theta(\Delta^2)$ 随着 n 增大，以高概率等于，因此我们希望在 $\Delta^2 = \widetilde{\Theta}(p/n)$. 处得到低阶 MMSE 下界。我们对 GAM 进行重新缩放，使先验分布满足我们的假设。

$Y = \sqrt{\lambda}X' + Z$，对于 $X' = X/\sqrt{\Delta}/s$ 和 $\lambda = \Delta/s.$。在这种缩放下， $\lambda_{\mathrm{ALG}} = \widetilde{\Theta}(\sqrt{p}/(s^2 n))$，我们的目标因此是~在 $\lambda = \Theta(\lambda_{\mathrm{ALG}})$.处证明低阶 MMSE 的困难性

事实证明，定理 2.7 适用于该稀疏聚类模型，因为此先验分布也满足假设 2.4，此处取 $D_n = \lfloor (1-\alpha)$对数 $n \rfloor$、威布尔常数 $\phi = 2$和 $A_n = \Theta($对数$^4\, n)$。特别地，定理 2.7推出以下结果。

定理 3.3。 设 p, n 为正整数，且 $p = n^{c+o(1)}$ 对某个常数 $c > 0$成立。对于上述稀疏聚类模型和先验分布，我们证明：当 $\lambda = \Delta/s$, $s = p^{a+o(1)}$ 时，其中 $a \in (1/2, 1)$；并且当 $D_n = \lfloor (1-\alpha)$对数 $n \rfloor$时：

$$\mathrm{MMSE}_X^{\leqslant D_n}\left(\frac{1}{\log^4 n}\cdot\left(\lambda + \frac{\mathrm{d}}{\mathrm{d}q}\mathcal{F}_{\mathrm{ann},\lambda}\Big|_{q=q(D_n)}\right)\right) = ns - \widetilde{O}\left(s\sqrt{n/p}\right) = (1+o(1))\mathrm{MMSE}_X^{\mathrm{trivial}}.$$

特别地，对于 $\lambda = \widetilde{\Theta}(\sqrt{p/(s^2 n)}) = \widetilde{\Theta}(\lambda_{\mathrm{ALG}})$,，我们有 $\mathrm{d}\,\frac{\mathrm{d}}{\mathrm{d}q}F_{\mathrm{ann},\,\lambda}\Big|_{q=q(D_n)} \geqslant 0$，因此

$$\mathrm{MMSE}_X^{\leqslant D_n}\left(\widetilde{\Theta}(\lambda_{\mathrm{ALG}})\right) \geqslant (1+o(1))\mathrm{MMSE}_X^{\mathrm{trivial}}.$$

证明推迟至第 10.3 节。

针对该模型中估计问题的这一紧致低阶下界，最近也已在 [EGV25b] 中得到证明。事实上，该工作证明了一个略强于我们的通用等价性所能推出的结果：通过利用聚类问题特有的潜变量结构，他们得到了 D 阶最小均方误差下界，其阶数可达到 D_n 略快于对数增长的程度。

4 结论

在这项工作中，我们为一大类高斯加性模型（GAM）严格地建立了退火 Franz-Parisi（FP）势的单调性与低阶 MMSE 下界之间的数学等价性。我们证明，退火 FP 势导数的符号，等价于低阶 MMSE 超过一个特定基准；该基准由从先验分布中独立同分布抽取的两个样本之间重叠的分位数所定义。我们的结果连接了这两种看似不同的计算困难性表述，它们不仅科学渊源不同，数学表述也各不相同。此外，我们的发现为未来研究开辟了若干有前景的方向：

- 退火势与淬火势： 我们的工作与经典统计物理学方法之间或许最显著的不同，在于我们关注的是退火 FP 势，而非淬火 FP 势。正如第 A 节所讨论的，这在某种程度上似乎是根本性的；对于某些 GAM，即使深入“低次易解”区域，淬火势仍然持续递增（表明处于“物理困难”区域）。相比之下，退火势的单调性精准地刻画了低阶 MMSE 从平凡转变为非平凡的相变。这表明，退火 FP 势或许更适合用于刻画统计问题中的计算困难性，也引出了一个根本问题：这一现象是否能在统计物理学中得到更深层次的解释。

- 势函数与算法类别之间的联系： 在估计任务中，淬火 FP 势与马尔可夫链蒙特卡罗方法的性能有关，这在一定程度上已是众所周知的结论，例如参见 [AGJ20、AWZ20]，其中证明了，淬火 FP 势的非单调性（即所谓的自由能垒）意味着马尔可夫链蒙特卡罗下界。此外，已知复本对称自由能势与秩一尖峰模型中近似消息传递（AMP）的性能有关（相关讨论参见 [MW24] ）。我们的工作（以及关于检测的 [BEAH+22、CMZZ25] ）建立了一种新的联系：退火 FP 势刻画了低次多项式的性能。进一步理解物理势函数与算法类别之间的这种映射，是一个非常有趣的未来研究方向。

- 从退火 FP 势中获取更多低次信息：我们证明，退火 FP 势的单调性决定了低阶 MMSE 是否优于某个特定

基准。在大多数应用中，这个基准的量级与“平凡均方误差”相同，也就是说，退火 FP 势递增意味着低次多项式无法超越平凡表现。虽然我们在应用中大量利用了这一结果的“单向”推论，但一个引人关注的开放问题是：退火势能否为低阶 MMSE 提供更精细的“成功”结果——具体而言，它能否准确指出低次多项式何时能达到优于平凡均方误差的低阶 MMSE的阶数。我们目前的结果表明，仅凭导数的符号不足以获取这一信息。

低阶互信息-均方误差关系：一种颇具说服力的方式来阐述我们方法的成功之处如下。信息论中最基本的恒等式之一是互信息-均方误差关系 [GSV05], ，它将任意 GAM 的互信息导数（即用物理学术语来说，系统“自由能”的导数）与最小均方误差联系起来。这一恒等式在刻画众多高维模型的渐近最小均方误差方面发挥了重要作用。具体而言，它是证明压缩感知 [BKM+19、RP16] 和低秩矩阵估计 [LM17], 的“复制对称”最小均方误差公式的关键步骤，也用于确立各种情形下的“全有或全无”相变 [RXZ19、NWZ20]。与直接分析最小均方误差相比，这种方法的技术优势在于，研究者可以转而研究更易处理的自由能，只需对其求导即可提取最小均方误差。

鉴于近期人们对低阶 MMSE 的兴趣激增——其分析通常采用直接组合方法和“累积量”界——自然会产生这样一个问题：是否存在低阶互信息-均方误差关系？理想情况下，这种关系可将低阶 MMSE 的分析简化为研究（希望更简单的）“低阶自由能”。我们的主要结果（定理 1.1）为一大类 GAM 确立了这样一个近似恒等式：我们证明，对于这些模型，退火 FP 势的导数与低阶 MMSE 相关联。由于退火 FP 势更容易计算，这使得一系列有趣模型的低阶 MMSE 分析相对容易。

由于经典的互信息-均方误差关系适用于所有 GAM，并可推广至泊松等其他情形 [AW12] 以及伯努利信道 [MNWS+23]，因此将我们的关键恒等式推广到高斯框架之外，是未来研究一个自然而有前景的方向。

5 证明的补充记号

我们记 $\mathbb{N}=\{1,\ 2,\ \ldots\}$和 $[N]=\{1,\ \ldots\ldots,\ N\}$。除非另有说明，所有渐近记号（$O,\ o,\ \Omega,\ \omega,\ \Theta$）均针对 $n\to\infty$，而 $\widetilde{O}(\cdot),\ \widetilde{\Omega}(\cdot),\ \widetilde{\Theta}(\cdot)$ 隐藏了多对数$(n)=(\text{对数}\,n)^{O(1)}$ 因子。对于一个事件 A，$\mathbf{1}_A$ 表示其示性函数。所有对数均为自然对数。

对于 $\alpha\in\mathbb{N}^N$，定义 $|\alpha|=\sum_{i=1}^N\alpha_i,\ \alpha!=\prod_{i=1}^N\alpha_i!$，对于 $X\in\mathbb{R}^N,\ X^\alpha=\prod_{i=1}^N X_i^{\alpha_i}$。我们用 $\alpha\geqslant\beta$表示逐坐标不等式，并在 $\alpha\geqslant\beta$时定义 $\binom{\alpha}{\beta}=\prod_{i=1}^N\binom{\alpha_i}{\beta_i}$。我们用 $\beta\prec\alpha$表示 $\beta\leqslant\alpha$和 $\beta\neq\alpha$。此外，$\mathrm{supp}(\alpha)=\{i\in[N]\colon\alpha_i\neq0\}$。

对于 $v\in\mathbb{R}^n$和 $r\geqslant1,\ v^{\otimes r}\in\mathbb{R}^{n^r}$表示 r 重张量幂，其中 $(v^{\otimes r})_{i_1\ldots i_r}=v_{i_1}\cdots v_{i_r}$.

对于两个随机变量 X,Y，我们用 $X\overset{d}{=}Y$表示分布相等，用 $X\perp\!\!\!\perp Y$表示独立性并且我们记 $X\sim P$ 表示随机变量 X 服从分布 P。

最后，对于任意 $D\in\mathbb{N}$，我们定义 $\exp_{\leqslant D}(v)=\sum_{k=0}^D\frac{v^k}{k!}$ 并且我们用 $\Gamma(t)$ 表示伽马函数，其定义为 $\Gamma(t)=\int_0^\infty x^{t-1}e^{-x}\mathrm{d}x,\ t>0$。

6 证明 FP 递增蕴含最小均方误差下界（定理2.7的证明）

6.1 累积量背景

我们介绍一些累积量的背景知识，以及它们满足的一些恒等式.

定义 6.1（累积量）。设 $X_1,\ldots,X_n$ 为联合分布随机变量。它们的累积量生成函数（CGF）是函数

$$K(t_1,\ldots,t_n)=\log\left(\mathbb{E}\left[\exp\left(\sum_{i=1}^n t_iX_i\right)\right]\right),$$

对所有使期望有限的 $t_1,\ldots,t_n\in\mathbb{R}$。联合累积量定义为

$$\kappa(X_1,\ldots,X_n)=\prod_{i=1}^n\frac{\mathrm{d}}{\mathrm{d}t_i}K(t_1,\ldots,t_n)\bigg|_{t_1=\ldots=t_n=0}.$$

利用多元泰勒-拉格朗日定理，对于使下列泰勒级数收敛的所有 $t=(t_1,\ldots,t_n)\in\mathbb{R}^n$，我们可以将其展开为级数：

$$K(t_1,\ldots,t_n)=\sum_{\alpha\in\mathbb{N}^n}^{+\infty}\frac{\prod_{i=1}^n\frac{\mathrm{d}}{\mathrm{d}t_i^{\alpha_i}}K(t_1,\ldots,t_n)\Big|_{t_1=\ldots=t_n=0}}{\alpha!}\prod_{i=1}^n t_i^{\alpha_i}.$$

以下命题的证明是标准的，参见例如 [SW22], 附录 D。

命题 6.2（独立性下消失）。设 a、$b\geqslant1$，且 X_1、…、X_a、Y_1、…、Y_b 为随机变量，并满足 $\{X_i\}_{i=1}^a$ 与 $\{Y_j\}_{j=1}^b$ 相互独立。则

$$\kappa(X_1,\ldots,X_a,Y_1,\ldots,Y_b)=0.$$

以下命题的证明推迟至 B 节。

命题 6.3（关于一个自变量的线性）。固定整数 $n\geqslant1,n\geqslant t\geqslant1$，并设 $X,Y,Z_1,\ldots,Z_{t-1},Z_{t+1},\ldots,Z_n$ 为联合分布的随机变量。那么，对于任意标量 $a,b\in\mathbb{R}$,

$$\begin{aligned}&\kappa\big(Z_1,\ldots,Z_{t-1},aX+bY,\,Z_{t+1},\ldots,Z_n\big)\\&\qquad=a\,\kappa\big(Z_1,\ldots,Z_{t-1},X,Z_{t+1},\ldots,Z_n\big)+b\,\kappa\big(Z_1,\ldots,Z_{t-1},Y,Z_{t+1},\ldots,Z_n\big).\end{aligned}$$

如下所示 累积量的标准公式。例如，可以在 [McC87], 公式 (2.9).

命题6.4 (矩-累积量分割公式)。 令 $P([n])$ 表示$[n] := \{1, \dots, n\}$的所有集合划分构成的集合。对于 $\pi \in P([n])$，记 $|\pi|$ 为 π中的块数。假设对于每个非空的 $B \subseteq [n]$，都有 $\mathbb{E}|X_B| < \infty$。联合累积量 $\kappa(X_1, \dots, X_n)$ 满足分拆公式

$$\kappa(X_1, \dots, X_n) = \sum_{\pi \in \mathcal{P}([n])} (|\pi| - 1)! \, (-1)^{|\pi|-1} \prod_{B \in \pi} \mathbb{E}\Big[\prod_{i \in B} X_i\Big]. \tag{6.1}$$

此外，矩满足矩—累积量关系：

$$\mathbb{E}\Big[\prod_{i=1}^{n} X_i\Big] = \sum_{\pi \in \mathcal{P}([n])} \prod_{B \in \pi} \kappa(X_i : i \in B). \tag{6.2}$$

以下命题的证明见 [SW22，附录 D]。

命题 6.5 (按拆分递归的累积量)。 设 $Y_1, \dots, Y_n$ 为随机变量，且对每个 $T \subseteq [n]$都有 $\mathbb{E}\big[\prod_{i \in T} |Y_i|\big] < \infty$。记 $\kappa(Y_i : \ i \in T)$ 为子族 $\{Y_i\}_{i \in T}$的联合累积量，并采用约定$\prod_{i \in} Y_i = 1$。则

$$\kappa(Y_1, \dots, Y_n) = \mathbb{E}\Big[\prod_{i=1}^{n} Y_i\Big] - \sum_{\varnothing \neq S \subseteq [n] \setminus \{1\}} \kappa\big(Y_i : i \notin S\big) \, \mathbb{E}\Big[\prod_{i \in S} Y_i\Big]. \tag{6.3}$$

引理 6.6 (对角切片恒等式)。 设 $X = \sum_{i=1}^{N} X_i$ ，并令 $\kappa_m(X)$ 为其第 m 阶累积量。则

$$\frac{\kappa_m(X)}{m!} = \sum_{|\gamma| = m} \frac{\kappa_\gamma(X_1, \dots X_N)}{\gamma!}. \tag{6.4}$$

证明。 根据定义，

$$\kappa_m(X) := \kappa(\underbrace{X, \dots, X}_{m \text{ times}}).$$

利用累积量的多重线性，即命题 6.3，以及定义 $X = \sum_{i=1}^{N} X_i$，

$$\kappa_m(X) = \kappa\Big(\sum_{i_1=1}^{N} X_{i_1}, \dots, \sum_{i_m=1}^{N} X_{i_m}\Big) = \sum_{i_1, \dots, i_m = 1}^{N} \kappa\big(X_{i_1}, \dots, X_{i_m}\big).$$

对于每个 m 元组 $(i_1, \dots, i_m)$，定义其重数向量

$$\gamma = (\gamma_1, \dots, \gamma_N), \qquad \gamma_k := \#\{j : i_j = k\}.$$

于是 $|\gamma| = \sum_{k=1}^{N} \gamma_k = m$。对于所有具有相同重数向量 γ 的 m 元组，由对称性可知，联合累积量 $\kappa(X_{i_1}, \dots, X_{i_m})$ 相同，且根据定义等于 κ_γ。这类 m 元组的数量，就是一个多重集中包含 γ_k 个 k 时的排列数，即 $\frac{m!}{\gamma!}$，其中 $\gamma! := \prod_{k=1}^{N} (\gamma_k)!$。因此，

$$\kappa_m(X) = \sum_{|\gamma| = m} \frac{m!}{\gamma!} \kappa_\gamma,$$

将其除以 $m!$，得到 (6.4)。 □

6.1.1 关键引理

为了证明低次相关性上界，我们需要一些关键引理；这些引理也将在论述过程中阐明证明思路。这些引理的证明推迟到第 6.3 节

首先，我们将用到一个简单的引理，它由一项重要结果推得；该结果通过先验分布的累积量给出了低阶相关性的上界 [SW22] 。具体而言，要得到低次相关性上界，只需对由此得到的累积量平方加权和进行界定。

引理 6.7（基于累积量的相关性界）。对于任意信噪比为 λ 的 *GAM*，

$$(\mathrm{Corr}_{P_0}^{\leqslant D})^2(\lambda) \leqslant \sum_{i=1}^{n} \sum_{0\leqslant|\alpha|\leqslant D} \frac{\kappa_\alpha^2(X_i, \sqrt{\lambda}X_1, \sqrt{\lambda}X_2, \ldots, \sqrt{\lambda}X_n)}{\alpha!},$$

其中，量 κ_α 在固定 $i \in [n]$时，等于以下相关随机变量集合的联合累积量：一个 X_i，以及对于每个 $j \in [n]$，有 α_j 个 X_j。

我们还将用到这个引理，它给出了高斯加性模型低阶 MMSE 的两种分解方式，我们会交替使用这两种方式。首先介绍一些符号说明。固定一个整数 $D \geqslant 0$，令 $\mathbf{R}[Y]_{\leqslant D}$ 表示由 Y 中总次数至多为 D的多项式构成的空间。将向量 X 的 D 次 MMSE 定义为：

$$\mathrm{MMSE}_X^{\leqslant D} := \inf_{f_1,\ldots,f_n \in \mathbf{R}[Y]_{\leqslant D}} \mathbb{E}\Big[\sum_{i=1}^{n} \big(f_i(Y) - X_i\big)^2\Big].$$

我们还分别定义所有 $i \in [n]$ 的 D 次逐坐标相关性，以及向量 X的相关性：

$$\mathrm{Corr}_{P_0,i}^{\leqslant D} := \sup_{\substack{g\in\mathbf{R}[Y]_{\leqslant D} \\ \mathbb{E}[g(Y)^2]=1}} \mathbb{E}\big[g(Y)X_i\big], \qquad \mathrm{Corr}_{P_0}^{\leqslant D} := \sup_{\substack{f_1,\ldots,f_n\in\mathbf{R}[Y]_{\leqslant D} \\ \mathbb{E}\sum_{i=1}^n f_i(Y)^2=1}} \mathbb{E}\Big[\sum_{i=1}^{n} f_i(Y)X_i\Big].$$

接下来，我们得到以下便于使用的公式。该引理的证明推迟到第B节。

引理 6.8（向量 X 的低阶 MMSE 公式）。假设我们有如上所述的高斯加性模型，其中 $X = (X_1, \ldots, X_n) \in \mathbf{R}^n$。那么：

$$\mathrm{MMSE}_X^{\leqslant D} = \sum_{i=1}^{n} \mathbb{E}[X_i^2] - \sum_{i=1}^{n} \big(\mathrm{Corr}_{P_0,i}^{\leqslant D}\big)^2 \tag{6.5}$$

$$= \mathbb{E}\|X\|^2 - \big(\mathrm{Corr}_{P_0}^{\leqslant D}\big)^2. \tag{6.6}$$

接下来，我们利用一个关于有界随机变量的对数矩母函数导数的简单结果。这将是对累积量平方之和进行界定的关键第一步。

引理 6.9 (截断对数矩母函数的有界导数)。设 $(\Omega, F, \mathbf{P})$ 是一个概率空间，Y_n 是一个实值随机变量，满足 $|Y_n| \leqslant M_n$，几乎处处成立，其中 $M_n > 0$。若 Y_n 的对数矩母函数 $\phi_n(\theta) :=$ 对数 $\mathbb{E}\big[e^{\theta Y_n}\big]$ 在开区间 $I \subset \mathbf{R}$ 上存在，则 $\phi_n(\theta)$ 在 I 上可微，且其导数满足

$$(\phi_n)'(\theta) \leqslant M_n, \qquad \textit{for all } \theta \in I.$$

为了应用上述引理，我们需要以下稍复杂的命题，它涉及随机变量在 0 附近的对数矩母函数泰勒展开及其与累积量的关系。将该命题与前面的引理结合起来，对于我们在证明中求得某些形式相似的累积量的加权和的界至关重要。

命题 6.10 (对数矩母函数的泰勒展开)。令 $T = \sum_{i=1}^n T_i$ 为 n 个随机变量 T_i 之和，满足 $|T| \leqslant M_n$，几乎处处成立，其中某个 $M_n > 0$。若 $\phi_n(\theta)$ 表示 T 的对数矩母函数，则 $\phi_n(\theta)$ 在 $\theta = 0$ 处的泰勒级数为

$$\phi_n(t) = \sum_{m\geqslant 1} \frac{\kappa_m(T)}{m!} t^m = \sum_{m\geqslant 1} t^m \sum_{|\gamma|=m} \frac{\kappa_{n,\gamma}(T_1,\ldots,T_n)}{\gamma!},$$

且该级数的收敛半径至少为 $R_n \geqslant 1/eM_n$. 此外，对于 $|t| < 1/(eM_n)$，次数为 m 的切片满足

$$\left|\sum_{|\gamma|=m} \frac{\kappa_{n,\gamma}}{\gamma!} t^{|\gamma|}\right| \leqslant \big(eM_n|t|\big)^m. \tag{6.7}$$

最后，接下来的两个引理使我们能够将前两个结果中的累积量界转化为所需的累积量平方和。为此，一个重要步骤是将截断随机变量的累积量与原始随机变量的累积量进行比较。以下引理保证了次 Weibull 随机变量的此类界。

引理 6.11。 固定 $m \geqslant 1$、指标 i_1、…、$i_m \in \{1$、…、$n\}$ 和 $M_n > 0$。设 T_i、$i \in [n]$ 为随机变量，满足对于某个 $\phi > 0$，T_i 对所有 $i \in [n]$ 都是次韦布尔(ϕ) 分布。若 $T = \sum_{i=1}^n T_i$、$T_i^{\mathrm{tr}} = T_i 1_{A_n}$ 以及 $\varepsilon_n = \mathrm{P}(|T| \geqslant M_n)$，则

$$\left|\kappa_m(T_{i_1}^{\mathrm{tr}}, \ldots, T_{i_m}^{\mathrm{tr}}) - \kappa_m(T_{i_1}, \ldots, T_{i_m})\right| \leqslant m^{C(\phi)m}\sqrt{\varepsilon_n}. \tag{6.8}$$

最后一步是将应用泰勒展开所产生的累积量（如命题 6.10 中所述）与 [SW22] 中得到的累积量平方进行比较。

我们引入一些符号说明。设 $X = (X_1$、…、$X_N) \in \mathrm{R}^N$ 为随机向量，并设 $X' = (X_1'$、…、$X_N')$ 为 X 的一个独立副本。对于多重指标 $\alpha = (\alpha_1$、…、$\alpha_N) \in \mathrm{N}^N$，我们定义

$$\widetilde{\kappa}_\alpha(X_1, \ldots, X_N) := \kappa_\alpha\big(X_1X_1', X_2X_2', \ldots, X_NX_N'\big). \tag{6.9}$$

引理 6.12。 固定 $\alpha \in \mathrm{N}^N$。假设对某些随机变量 $X_1, \ldots, X_N$，对于任意基数为 $|\alpha| \leqslant D_n$ 的多重指标 $\alpha \in \mathrm{N}^N$，都有：

$$\kappa_\alpha(X_1, \ldots, X_N) \geqslant 0. \tag{6.10}$$

于是，我们可以证明如下结论：对于所有满足 $|\alpha| \leqslant D_n$ 的 $\alpha \in \mathrm{N}^N$，都有：

$$\widetilde{\kappa}_\alpha(X_1, \ldots, X_N) \geqslant \kappa_\alpha^2(X_1, \ldots, X_N).$$

6.2 上界证明

定理 2.7 的证明。 记 ϕ_n 为 $\langle X, X'\rangle$ 的对数矩母函数，其中 X, X' 是从 P_0 中独立同分布抽取的两个样本。我们还将“截断”随机变量 $\langle X, X'\rangle 1_{(|\langle X, X'\rangle| \leqslant q(D_n'))}$ 的对数矩母函数记为 ϕ_n^{tr}。

接着，注意到应用引理 6.9 于 $M_n = q(D_n')$，可得对于所有 $\theta \in I := \{t \in \mathrm{R} : \phi_n^{\mathrm{tr}}(t) < \infty\}$，都有：

$$(\phi_n^{\mathrm{tr}})'(\theta) \leqslant q(D_n'). \tag{6.11}$$

设 R_n 为 ϕ_n^{tr} 在 0 附近的泰勒展开半径。根据命题 6.10，$(q(D_n'))^{-1} \leqslant R_n$ 因此，根据同一命题，对于所有 $|\theta| \leqslant (q(D_n'))^{-1}$：

$$\phi_n^{tr}(\theta) = \sum_{m=1}^{\infty} \theta^m \sum_{|\gamma|=m} \frac{\kappa_{n,\gamma}^{\mathrm{tr}}(X_1X_1', \ldots, X_nX_n')}{\gamma!}.$$

求导

$$(\phi_n^{tr})'(\theta) = \sum_{m=1}^{\infty} m\theta^{m-1} \sum_{|\gamma|=m} \frac{\kappa_{n,\gamma}^{tr}(X_1X_1', \ldots, X_nX_n')}{\gamma!}.$$

现在，再次利用命题 6.10，若 $r_n =$ 对数 n，则对于 $|\theta| \leqslant (q(D_n')r_n)^{-1}$，有 $q := q(D_n')|\theta| \leqslant 1/r_n$，而次数超过 $D_n + 1$ 的尾项满足：

$$\left|\sum_{m \geqslant D_n+2} mt^{m-1} \sum_{|\gamma|=m} \frac{\kappa_{n,\gamma}^{tr}}{\gamma!}\right| \leqslant \sum_{m \geqslant D_n+2} (q(D_n')|\theta|)^m \leqslant \frac{q^{D_n+2}}{1-q} \leqslant \frac{(1/r_n)^{D_n+2}}{1-1/r_n}.$$

因此，对于所有 $|\theta| \leqslant (q(D_n')r_n)^{-1}$，我们得出：

$$\left|(\phi_n^{tr})'(\theta) - \sum_{m=1}^{D_n+1} m\theta^{m-1} \sum_{|\gamma|=m} \frac{\kappa_{n,\gamma}^{tr}(X_1X_1', \ldots, X_nX_n')}{\gamma!}\right| \leqslant \frac{(1/r_n)^{D_n+2}}{1-1/r_n}. \tag{6.12}$$

结合6.11) 可推出

$$\sum_{m=1}^{D_n+1} m\theta^{m-1} \sum_{|\gamma|=m} \frac{\kappa_{n,\gamma}^{tr}(X_1X_1', \ldots, X_nX_n')}{\gamma!} - \frac{(1/r_n)^{D_n+2}}{1-1/r_n} \leqslant q(D_n'). \tag{6.13}$$

但是，利用引理 6.11 可得，对所有满足 γ 的 $|\gamma| \leqslant D_n$

$$\left|\kappa_\gamma^{\mathrm{tr}}(X_1, \ldots, X_n) - \kappa_\gamma(X_1, \ldots, X_n)\right| \leqslant |\gamma|^{C(\phi)|\gamma|} e^{-D_n'} \leqslant e^{C(\phi)D_n \log D_n - D_n'}, \tag{6.14}$$

对于某个 $C(\phi) > 0$，其取值取决于假设 $\phi > 0$ 中第 3 项的次威布尔常数 2.4。最后，注意到有 n^{D_n+1} 个这样的多重指标 γ，使得 $|\gamma| \leqslant D_n + 1$。将（6.14）应用于每一项，经直接代数运算可得，

$$\begin{aligned}
&\sum_{|\gamma|\leqslant D_n+1} \frac{\kappa_{n,\gamma}^{tr}(X_1X_1', \ldots, X_nX_n') - \kappa_{n,\gamma}(X_1X_1', \ldots, X_nX_n')}{\gamma!} |\gamma|\theta^{|\gamma|-1} \\
&\quad \geqslant -e^{C(\phi)D_n \log D_n - D_n'} \cdot \sum_{|\gamma|\leqslant D_n+1} |\gamma|\theta^{|\gamma|-1}(\gamma!)^{-1} \\
&\quad \geqslant -e^{C(\phi)D_n \log D_n - D_n'} \cdot \sum_{0<|\gamma|\leqslant D_n+1} \theta^{|\gamma|-1} \\
&\quad \geqslant -e^{C(\phi)D_n \log D_n + D_n \log n - D_n'} \cdot \max\{\theta, \theta^{D_n}\}.
\end{aligned}$$

进一步假设 $0 < \theta < n^{C_1}$，其中 $C_1 > 0$ 是某个正的普适常数，则可得对所有 $|\theta| \leqslant \min\{(q(D_n')r_n)^{-1},\ n^{C_1}\}$,

$$\sum_{|\gamma|\leqslant D_n+1} \frac{\kappa_{n,\gamma}^{tr}(X_1X_1', \ldots, X_nX_n') - \kappa_{n,\gamma}(X_1X_1', \ldots, X_nX_n')}{\gamma!} |\gamma|\theta^{|\gamma|-1} \geqslant -e^{C(\phi)D_n \log D_n + (C_1+1)D_n \log n - D_n'} \tag{6.15}$$

因此，（6.13）与（6.15）共同推出，对所有 $|\theta| \leqslant \min\{(q(D_n')r_n)^{-1},\ n^{C_1}\}$,

$$\sum_{m=1}^{D_n+1} m\theta^{m-1} \sum_{|\gamma|=m} \frac{\kappa_{n,\gamma}(X_1X_1', \ldots, X_nX_n')}{\gamma!} - \frac{(1/r_n)^{D_n+2}}{1-1/r_n} - e^{C(\phi)D_n \log D_n + (C_1+1)D_n \log n - D_n'} \leqslant q(D_n'). \tag{6.16}$$

但现在请注意，我们关于信号的先验分布 X 满足假设2.4 中的第 2 项，因此，引理6.12结合初等代数运算可得，

$$\begin{aligned}
\sum_{m=1}^{D_n+1} m\theta^{m-1} \sum_{|\gamma|=m} \frac{\kappa_{n,\gamma}(X_1X_1', \ldots, X_nX_n')}{\gamma!} &= \sum_{|\gamma|\leqslant D_n+1} \frac{\kappa_{n,\gamma}(X_1X_1', \ldots, X_nX_n')}{\gamma!} |\gamma|\theta^{|\gamma|-1} && (6.17) \\
&\geqslant \sum_{|\gamma|\leqslant D_n+1} \frac{\kappa_{n,\gamma}^2(X_1, \ldots, X_n)}{\gamma!} |\gamma|\theta^{|\gamma|-1} && (6.18) \\
&= \sum_{i=1}^n \sum_{|\gamma|\leqslant D_n} \frac{\kappa_{n,\gamma}^2(X_i, \sqrt{\theta}X_1 \ldots, \sqrt{\theta}X_n)}{(\gamma+e_i)!} (|\gamma|+1) && (6.19) \\
&\geqslant \sum_{i=1}^n \sum_{|\gamma|\leqslant D_n} \frac{\kappa_{n,\gamma}^2(X_i, \sqrt{\theta}X_1 \ldots, \sqrt{\theta}X_n)}{\gamma!}, && (6.20)
\end{aligned}$$

其中，在最后一个不等式中，我们用到了对所有 $i \in \mathbf{N}$， $\gamma!(|\gamma|+1) \geqslant (\gamma+e_i)!$。

现在，根据定理 6.7，对每个 $i=1,\dots,n$，

$$(\mathrm{Corr}_{P_0,i}^{\leqslant D})^2(\theta)_{\leqslant D_n} \leqslant \sum_{|\gamma|\leqslant D_n} \frac{\kappa_{n,\gamma}^2(X_i,\sqrt{\theta}X_1\dots,\sqrt{\theta}X_n)}{\gamma!}.$$

对 i求和，然后将最后显示的不等式与（6.20）和（6.16结合起来，我们得出，对于所有 $|\theta| \leqslant \min\{(q(D_n')r_n)^{-1}, n^{C_1}\}$,

$$(\mathrm{Corr}_{P_0}^{\leqslant D})^2(\theta) - \frac{(1/r_n)^{D_n+2}}{1-1/r_n} - e^{C(\phi)D_n\log D_n+(C_1+1)D_n\log n-D_n'} \leqslant q(D_n'). \tag{6.21}$$

使用它 根据假设 2.4中的第 3 项，代入 $r_n =$ 对数 n，并取 n 足够大，我们知道 t:

$$e^{(C(\phi)+C+1)D_n\log(nD_n)-D_n'} \leqslant e^{-D_n'/2} \leqslant n^{-\frac{D_n}{2}\log n} \leqslant (\log n)^{-D_n/2} = o(q(D_n'))$$
$$\frac{(1/r_n)^{D_n+2}}{1-1/r_n} \leqslant 2(\log n)^{-D_n} = o(q(D_n')).$$

因此，对于所有 $|\theta| \leqslant \min\{(q(D_n')r_n)^{-1}, n^{C_1}\}$，如果 n 足够大，

$$(\mathrm{Corr}_{P_0}^{\leqslant D})^2(\theta) \leqslant 2q(D_n'). \tag{6.22}$$

但是，现在注意到恒等式：对所有 $u>0$,

$$u + \frac{d}{dq}\mathcal{F}_{\mathrm{ann},u}|_{q=q(D_n)} = -\frac{d}{dq}\log \mathbb{P}(\langle X,X'\rangle = q)|_{q=q(D_n)}.$$

结合假设中的条件 2 2.4可得

$$\frac{1}{A_n}\left(u + \frac{d}{dq}\mathcal{F}_{\mathrm{ann},u}|_{q=q(D_n)}\right) \leqslant (q(D_n')r_n)^{-1}.$$

接下来注意到，利用假设中的第 3 项 2.4，该量有如下上界： $n^C/$对数 n其中 $C>0$是某个普适常数。因此，我们可以取 $C_1 = C$，然后将 $\theta = \frac{1}{A_n}\left(u + \frac{d}{dq}\mathcal{F}_{\mathrm{ann},\ u}|_{q=q(D_n)}\right) \leqslant n^{C_1}$ 代入 （6.22）→，从而完成证明。 □

6.3 关键引理的证明

6.3.1 引理 6.9 的证明

由于 Y_n 有界且上界为 M_n,

$$|Y_n(\omega)| \leqslant M_n \qquad \text{for all } \omega \in \Omega.$$

接下来， $\phi_n(\theta)$ 在 I 上有限，且 Y_n 有界。根据控制收敛定理， ϕ_n 在I 上可微，并且对每个 $\theta \in I$，都有

$$(\phi_n)'(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}\log \mathbb{E}\left[e^{\theta Y_n}\right] = \mathbb{E}_\theta\left[Y_n\right],$$

其中 $\mathbb{E}_\theta$ 表示相对于由以下定义的倾斜测度 P_θ 的期望

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}} = \frac{e^{\theta Y_n^{\mathrm{tr}}}}{\mathbb{E}\left[e^{\theta Y_n^{\mathrm{tr}}}\right]}.$$

利用 Y_n 上的逐点界以及 P_θ 是概率测度这一事实，我们得到

$$\left|(\phi_n)'(\theta)\right| = \left|\mathbb{E}_\theta[Y_n]\right| \leqslant \mathbb{E}_\theta\left[|Y_n|\right] \leqslant M_n,$$

对所有 $\theta \in I$ 均成立，正如所述。

6.3.2 命题的证明6.10

在继续之前　　证明过程中，我们陈述如下引理，其证明推迟至第6.4节。

引理 6.13 （有界变量的累积量界）。 设 Z 是一个实值随机变量，满足 $|Z| \leqslant B$ 几乎处处成立，其中 $B > 0$。令 $\kappa_m(Z)$ 表示其 m 阶累积量，其中 $m \geqslant 1$。那么

$$|\kappa_m(Z)| \leqslant m!(eB)^m, \qquad m = 1, 2, \ldots \tag{6.23}$$

我们现在结合引理 6.6 和引理 6.13 来证明命题 6.10。

命题 *6.10* 的证明。 由于 $|T| \leqslant M_n$，矩母函数（MGF）

$$\Phi_n(t) := \mathbb{E}e^{tT}$$

对所有 $t \in \mathbf{R}$ 都是解析函数。此外， $\Phi_n(0) = 1$，因此当 $|t|$ 足够小时， $\Phi_n(t) \neq 0$ ，故 T 的对数矩母函数

$$\phi_n(t) := \log \Phi_n(t)$$

在原点的某个邻域内解析。它在 0 处的泰勒展开为

$$\phi_n(t) = \sum_{m \geqslant 1} \frac{\phi_n^{(m)}(0)}{m!} t^m = \sum_{m \geqslant 1} \frac{\kappa_m(T)}{m!} t^m,$$

其中 $\phi_n^{(m)}(0)$ 表示 m 阶导数，并等于 m 的第 T 个累积量。由引理 6.13 应用于$Z = Y$和 $B = M_n$，可得

$$\left|\kappa_m(T)\right| \leqslant m!\ (eM_n)^m.$$

因此，泰勒系数 $a_m := \kappa_m(T)/m!$ 满足 $|a_m| \leqslant (eM_n)^m$. 由根值判别法，

$$\limsup_{m \to \infty} |a_m|^{1/m} \leqslant eM_n,$$

因此，泰勒级数的收敛半径 R_n 至少为

$$R_n \geqslant \frac{1}{eM_n}.$$

特别地，该级数对所有 $|t| < 1/(eM_n)$ 都绝对收敛。现在考虑以下级数

$$\sum_{\gamma \neq 0} \frac{\kappa_{n,\gamma}(T_1, \ldots, T_n)}{\gamma!} t^{|\gamma|} = \sum_{m \geqslant 1} t^m \sum_{|\gamma| = m} \frac{\kappa_{n,\gamma}(T_1, \ldots, T_n)}{\gamma!}.$$

由引理 6.6，系数 t^m 等于

$$\sum_{|\gamma| = m} \frac{\kappa_{n,\gamma}}{\gamma!} = \frac{\kappa_m(T)}{m!}.$$

因此

$$\phi_n(t) = \sum_{m \geqslant 1} \frac{\kappa_m(T)}{m!} t^m = \sum_{m \geqslant 1} t^m \sum_{|\gamma| = m} \frac{\kappa_{n,\gamma}}{\gamma!}.$$

此外，由引理6.13，

$$\left| \sum_{|\gamma| = m} \frac{\kappa_{n,\gamma}}{\gamma!} t^{|\gamma|} \right| = \left| \frac{\kappa_m(T)}{m!} t^m \right| \leqslant (eM_n)^m |t|^m = (eM_n|t|)^m,$$

也就是（6.7），从而完成证明。 □

6.3.3 引理6.11的证明

本节旨在证明引理 6.11。为此，我们先介绍一些符号，并证明两个有助于此的结果（引理 6.14 和 6.15）。固定一个整数 $n \geqslant 1$，令 T_1、…、T_n 为定义在同一概率空间上的任意实值随机变量，并定义 $T := \sum_{i=1}^n T_i$. 固定 $m \geqslant 1$ 及指标i_1、…、$i_m \in \{1$、…、$N\}$。我们关注随机变量 m的第阶联合累积量

$$T_{i_1}, \dots, T_{i_m} \quad \text{and} \quad T_{i_1}^{\mathrm{tr}}, \dots, T_{i_m}^{\mathrm{tr}}, \qquad \text{for all} \qquad i_1, \dots, i_m \in [N]$$

并界定它们的差值。为方便起见，定义

$$L := \max_{1 \leqslant t \leqslant m} \left(\mathbb{E}|T_{i_t}|^m\right)^{1/m}, \qquad U := \max_{1 \leqslant t \leqslant m} \left(\mathbb{E}|T_{i_t}|^{2m}\right)^{1/(2m)}. \tag{6.24}$$

为简化符号说明，我们令 $Z_t := T_{i_t}$, $Z_t^{\mathrm{tr}} := T_{i_t}^{\mathrm{tr}} = T_{i_t} 1_{A_n}$, $t = 1, \dots, m$, 从而

$$\kappa_m(T_{i_1}, \dots, T_{i_m}) = \kappa_m(Z_1, \dots, Z_m), \qquad \text{and} \qquad \kappa_m(T_{i_1}^{\mathrm{tr}}, \dots, T_{i_m}^{\mathrm{tr}}) = \kappa_m(Z_1^{\mathrm{tr}}, \dots, Z_m^{\mathrm{tr}}).$$

此外， L, U 现在将为:

$$L := \max_{1 \leqslant t \leqslant m} \left(\mathbb{E}|Z_t|^m\right)^{1/m}, \qquad U := \max_{1 \leqslant t \leqslant m} \left(\mathbb{E}|Z_t|^{2m}\right)^{1/(2m)}, \tag{6.25}$$

假设对所有 t， $\mathbb{E}|Z_t|^{2m} < \infty$ 均成立，从而一切都有定义。接下来，给定一个块 $B \subseteq \{1, \dots, m\}$，定义

$$Z_B := \prod_{t \in B} Z_t, \qquad s := |B|.$$

我们将使用以下引理，其证明推迟到第 6.4节。

引理 6.14 (在 A_n 下的截断块界)。 设 $B \subseteq \{1, \dots, m\}$ 是一个大小为 $s = |B|$ 的块。令L, U 如 (6.24) 和 $\varepsilon_n = \mathbb{P}(A_n^c)$中所定义，则有

$$\mathbb{E}\big|Z_B\big| \leqslant L^s \qquad \textit{and} \qquad \mathbb{E}\big[|Z_B| 1_{A_n^c}\big] \leqslant U^s \sqrt{\varepsilon_n}. \tag{6.26}$$

利用这个引理，我们现在可以证明以下结果，它将截断随机变量的多元累积量与原始随机变量的多元累积量联系起来。

引理 6.15 (截断下的累积量)。 固定 $m \geqslant 1$ 和指标 $i_1, \dots, i_m \in \{1, \dots, N\}$。假设$\mathbb{E}|T_{i_t}|^{2m} < \infty$ 对所有 t均成立。设 L, U 由 (6.24) 和上文中的 $\varepsilon_n = \mathbb{P}(|Y| \geqslant M_n)$ 给出。那么，

$$\left|\kappa_m(Z_1^{\mathrm{tr}}, \dots, Z_m^{\mathrm{tr}}) - \kappa_m(Z_1, \dots, Z_m)\right| \leqslant (m-1)!\, m^m\, 2^m\, (U \vee L)^m\, \sqrt{\varepsilon_n}. \tag{6.27}$$

证明。 回顾分 割公式 (6.1) 应用于截断变量 $Z_t^{\mathrm{tr}} = X_{i_t}^{\mathrm{tr}}$和原始变量 s $Z_t = X_{i_t}$:

$$\kappa_m(Z_1^{\mathrm{tr}}, \dots, Z_m^{\mathrm{tr}}) - \kappa_m(Z_1, \dots, Z_m) = \sum_{\pi \in \Pi_m} (|\pi| - 1)!(-1)^{|\pi|-1} \Delta_\pi,$$

其中

$$\Delta_\pi := \prod_{B \in \pi} \mathbb{E}\Big[\prod_{t \in B} Z_t^{\mathrm{tr}}\Big] - \prod_{B \in \pi} \mathbb{E}\Big[\prod_{t \in B} Z_t\Big].$$

固定 $\pi \in \Pi_m$。对于一个块 $B \in \pi$，我们记 $Z_B := \prod_{t \in B} Z_t$。那么

$$\mathbb{E}\Big[\prod_{t \in B} Z_t^{\mathrm{tr}}\Big] = \mathbb{E}\big[Z_B 1_{A_n}\big] = \mathbb{E}[Z_B] - \mathbb{E}\big[Z_B 1_{A_n^c}\big].$$

因此，

$$\prod_{B \in \pi} \mathbb{E}\Big[\prod_{t \in B} Z_t^{\mathrm{tr}}\Big] = \prod_{B \in \pi} \Big(\mathbb{E}[Z_B] - \mathbb{E}\big[Z_B 1_{A_n^c}\big]\Big).$$

减去 $\prod_{B\in\pi}\mathbb{E}[Z_B]$ 并展开乘积，可得

$$\Delta_\pi = \sum_{\varnothing\neq S\subseteq\pi}(-1)^{|S|}\prod_{B\in S}\mathbb{E}\big[Z_B 1_{A_n^c}\big]\prod_{B\notin S}\mathbb{E}[Z_B].$$

取绝对值并应用引理 6.14 可得，对于每个非空的 $S\subseteq\pi$，

$$\prod_{B\in S}\big|\mathbb{E}[Z_B 1_{A_n^c}]\big|\prod_{B\notin S}\big|\mathbb{E}[Z_B]\big| \leqslant \prod_{B\in S}\mathbb{E}\big[|Z_B|1_{A_n^c}\big]\prod_{B\notin S}\mathbb{E}|Z_B| \leqslant \prod_{B\in S}U^{|B|}\sqrt{\varepsilon_n}\prod_{B\notin S}L^{|B|}.$$

令 $r(S):=\sum_{B\in S}|B|$，并注意到 $\sum_{B\in\pi}|B|=m$。于是

$$\prod_{B\in S}U^{|B|}\sqrt{\varepsilon_n}\prod_{B\notin S}L^{|B|} = (\sqrt{\varepsilon_n})^{|S|}\,U^{r(S)}\,L^{m-r(S)} \leqslant \sqrt{\varepsilon_n}\,(U\vee L)^m,$$

因为 $|S|\geqslant 1$ 蕴含 $(\sqrt{\varepsilon_n})^{|S|}\leqslant\sqrt{\varepsilon_n}$（我们总有 $\varepsilon_n\leqslant 1$），且 $U^rL^{m-r}\leqslant(U\vee L)^m$。因此，

$$|\Delta_\pi| \leqslant \sum_{\varnothing\neq S\subseteq\pi}\sqrt{\varepsilon_n}\,(U\vee L)^m = (2^{|\pi|}-1)\,(U\vee L)^m\sqrt{\varepsilon_n} \leqslant 2^m\,(U\vee L)^m\sqrt{\varepsilon_n},$$

using $|\pi|\leqslant m$ 对 $[m]$ 的任意划分 π 都成立。将此估计代入划分公式 a,

$$\big|\kappa_m(Z_1^{\mathrm{tr}},\dots,Z_m^{\mathrm{tr}})-\kappa_m(Z_1,\dots,Z_m)\big| \leqslant \sum_{\pi\in\Pi_m}(|\pi|-1)!\,|\Delta_\pi| \leqslant 2^m(U\vee L)^m\sqrt{\varepsilon_n}\sum_{\pi\in\Pi_m}(|\pi|-1)!,$$

这几乎就得到了我们的第一个不等式。为了得到 (6.27)，注意到 m^m 的集合划分至多有{1、.....、 m}，并且对于所有 $\pi\in\Pi_m$，都有 $(|\pi|-1)!\leqslant(m-1)!$，因此

$$\sum_{\pi\in\Pi_m}(|\pi|-1)!\leqslant(m-1)!m^m.$$

$\sum_{\pi\in\Pi_m}(|\pi|-1)!\leqslant(m-1)!\,m^m$，因此

$$\big|\kappa_m(Z_1^{\mathrm{tr}},\dots,Z_m^{\mathrm{tr}})-\kappa_m(Z_1,\dots,Z_m)\big| \leqslant (m-1)!\,m^m\,2^m\,(U\vee L)^m\,\sqrt{\varepsilon_n}. \tag{6.28}$$

这就是我们所需的不等式，证明完毕。 □

如果我们希望得到一个对索引 $(i_1、.....、 i_m)$ 的选择均成立的界，只需将（6.24）替换为

$$L_n := \max_{1\leqslant i\leqslant n}\big(\mathbb{E}|T_i|^m\big)^{1/m}, \qquad U_n := \max_{1\leqslant i\leqslant n}\big(\mathbb{E}|T_i|^{2m}\big)^{1/(2m)}. \tag{6.29}$$

那么，对于所有 $L\leqslant L_n$ 和 $U\leqslant U_n$，$(i_1、.....、 i_m)$，引理 6.15 在将 L、U 替换为 L_n、U_n 后仍然成立。

现在我们可以证明主定理所需的引理 6.11 了。

引理 *6.11* 的证明。 注意， $2^m m!\leqslant m^{m\log 2}$ 且 $\sqrt{\varepsilon}\leqslant 1$ 显然成立。因此，要使我们的结论成立，只需证明

$$\max\{L_n,U_n\}\leqslant m^{c_1(\phi)m},$$

对所有充分大的 n 以及常数 $c_1(\phi)>0$，其中 ϕ 是 X_i 的次韦布尔常数。我们将证明，所有满足假设 2.4 的先验分布都满足这一点。事实上，假设 2.4 的第 3 项意味着所有坐标 $(T_i)_{i=1}^N$ 均为一致次韦布尔($\phi/2$)（因为 $T_i=X_iX_i'$）。 T_i 为次韦布尔($\phi/2$) 对某个 $\phi>0$ 成立，这根据定义 2.2等价于：存在一个与 K 和 i 独立的常数 n:

$$(\mathbb{E}|T_i|^p)^{1/p}\leqslant C_\phi K p^{2/\phi} \qquad \text{for all } p\geqslant 1,$$

因此

$$L\leqslant C_\phi K m^{2/\phi}, \qquad U\leqslant C_\phi K(2m)^{2/\phi}.$$

因此 $\max\{U_n,\ L_n\}^m\leqslant(2^{1/\phi}C_\phi K)^m m^{2m/\phi}$，我们的证明也就完成了。

□

6.3.4 引理6.12 的证明

本节旨在证明引理 6.12。我们先介绍一些符号说明和一个定义。对于 N 随机变量 X_1，……， X_N我们曾记作

$$\tilde{\kappa}_\alpha(X_1,\dots,X_n) := \kappa_\alpha(X_1X_1',\dots,X_nX_n'),$$

其中，向量 $X' = (X_1',\dots,X_n')$ 是 $X = (X_1,\dots,X_n)$的一个独立同分布副本。本小节将给出我们上文定义的累积量 $\tilde{\kappa}_\alpha$ 的闭式表达。我们的结果基于 [McC87]的一项重要结果。我们先给出以下定义：

定义 6.16. 对于任意划分 $\pi \in \mathcal{P}([n])$，我们将由该划分π“诱导”出的图记为 $E(\pi)$，即顶点 $[n]$, 上的图 G，它是 $|\pi|$ 个团 $(C_S)_{S\in\pi}$的并，即 $G = \bigcup_{S\in\pi} C_S$, 其中，对于划分 $S \in \pi$ 的每个部分，我们将顶点 $i \in S$上的团记为 C_S。

定理 6.17 (等式 (3.3)，引自 [McC87])。 设有 n 个随机变量 $X_1,\dots,X_n$。那么，对于任意划分 $\pi \in \mathcal{P}([n])$, 其部分为 $S_1,\dots,S_{|\pi|}$，以下结论成立：

$$\kappa(\prod_{i\in S_1} X_i,\dots,\prod_{i\in S_{|\pi|}} X_i) = \sum_{\pi'\in G}\prod_{S'\in\pi'}\kappa(S'). \tag{6.30}$$

其中 $G := \{\pi' \in P([n])$ 使得 $E(\pi)\cup E(\pi')$ 是一个连通图 $\}$。

利用这一结果，我们可以证明感兴趣的累积量 $\tilde{\kappa}$ 满足以下结果。对任意 $N \in \mathbf{N}$，定义划分 $\pi^N_{\text{pair}} = \{\{1,\ 2\},\ \{3,\ 4\},\ \dots\dots,\ \{2N-1,\ 2N\}\}$。

引理 6.18。 设 $X_1,\dots,X_N$ 是一列同分布但不一定相互独立的随机变量。任取 $\alpha \in \mathbf{N}^N$，并设 $X' = (X_1',\dots,X_N')$ 是 $X = (X_1,\dots,X_N)$ 的独立同分布副本。记 $\widetilde{X} \in \mathbf{R}^{2|\alpha|}$ 为如下定义的向量

$$\widetilde{X} := (\underbrace{X_1,X_1'\dots,X_1,X_1'}_{\alpha_1-times},\dots,\underbrace{X_N,X_N'\dots,X_N,X_N'}_{\alpha_N-times}).$$

那么，它们的联合累积量 $\tilde{\kappa}_\alpha := \kappa_\alpha(X_1X_1',\dots,X_NX_N')$ 满足：

$$\tilde{\kappa}_\alpha := \kappa_\alpha(X_1X_1',\dots,X_NX_N') = \kappa(\widetilde{X}_1\widetilde{X}_2,\dots,\widetilde{X}_{2|\alpha|-1}\widetilde{X}_{2|\alpha|}) = \sum_{(\pi_1,\pi_2)\in J}\prod_{S\in\pi_1,\ S\in\pi_2}\kappa(S)\kappa(S') \tag{6.31}$$

其中 $J := \{(\pi_1,\pi_2) \in P(\{1,3,\dots,2|\alpha|-1\})\times P(\{2,,4,\dots,2|\alpha|\})$ 使得 $E(\pi_1)\cup E(\pi_2)\cup E(\pi^{|\alpha|}_{\text{划分对}})$ 是连通图$\}$。

证明。应用定理 6.17，对于 $\pi^{|\alpha|}_{\text{划分对}} = \{\{1,\ 2\},\ \{3,\ 4\},\ \dots\dots,\ \{2|\alpha|-1,\ 2|\alpha|\}\}$ 关于坐标~的 X 划分，我们有

$$\tilde{\kappa}_\alpha := \kappa_\alpha(X_1X_1',\dots,X_NX_N') = \sum_{\pi'\in G}\prod_{S'\in\pi'}\kappa(S')$$

其中 G:= $\{\pi' \in \mathcal{P}([2|\alpha|])$ 使得 $E(\pi^{|\alpha|}_{\text{划分对}})\cup E(\pi')$ 是连通图$\}$。换言之，第一个求和遍历所有划分 π'，使得它们与划分 $\pi^{|\alpha|}_{\text{划分对}}$组合后构成连通图。

我们的证明只需完成以下两个步骤即可。首先，我们证明 J 中的每个划分对都会出现在 G所定义的求和中。其次，我们说明没有其他划分会对求和产生非零贡献。

对于第一部分，令 $(\pi_1,\ \pi_2) \in J$。由于 π_1 划分奇数指标（来自向量 X），且 π_2划分偶数指标（来自向量 X'），它们的并集 $\pi = \pi_1\cup\pi_2$ 是 $[2|\alpha|]$上的一个划分。此外，由于 $E(\pi) = E(\pi_1)\cup E(\pi_2)$、 π 属于 G，这一步得以完成。

对于第二步，考虑一个划分 π^* 不属于 J 但属于 G的划分。于是 π^* 无法分解为关于 X 变量和 X' 变量的两个独立划分。接下来我们论证 $\prod_{S'\in\pi^*}\kappa(S') = 0$。确实，在这种情况下， π^* 包含一个块 S^*以及指标 i、 j，使得 $X_i \in S^*$ 且 $X_j' \in S^*$。但是向量 X、 X' 相互独立，因此由命题 6.2可得

□

现在我们可以证明引理6.12。

引理 *6.12*的证明。我们使用关于 6.31的公式（$\tilde{\kappa}$），并注意到其中两点。首先，如果 $\psi = \{1, 2, \ldots\ldots, |\alpha|\}$那么 $(\psi, \psi) \in J$。因此，

$$\tilde{\kappa}_\alpha(X_1, \ldots, X_N) = \kappa_\alpha^2(X_1, \ldots, X_N) + \sum_{(\pi_1,\pi_2)\in J\setminus\{(\psi,\psi)\}} \prod_{S\in\pi_1,\ S\in\pi_2} \kappa(S)\kappa(S').$$

此外，剩余总和中的任何其他项都是一个乘积，其中只包含随机变量 X_1、……、X_N 的联合累积量（可能重复出现）；回忆一下，X' 是独立同分布的，且 X）。然而，根据我们推论中的假设，所有此类累积量都是非负的，因此，

$$\tilde{\kappa}_\alpha(X_1, \ldots, X_N) \geqslant \kappa_\alpha^2(X_1, \ldots, X_N).$$

□

6.4 辅助引理的证明

在本节中，我们证明引理6.13和6.14。

引理 *6.13*的证明。 回顾单个变量的矩—累积量公式 6.4:

$$\kappa_m(Z) = \sum_{\pi\in\mathcal{P}([m])} (|\pi| - 1)!(-1)^{|\pi|-1} \prod_{B\in\pi} \mathbb{E}\Big[Z^{|B|}\Big].$$

取绝对值并利用 $|Z| \leqslant B$,

$$|\kappa_m(Z)| \leqslant \sum_{\pi\in\mathcal{P}([m])} (|\pi| - 1)! \prod_{B\in\pi} \mathbb{E}|Z|^{|B|} \leqslant \sum_{\pi\in\mathcal{P}([m])} (|\pi| - 1)! \prod_{B\in\pi} B^{|B|} = B^m \sum_{\pi\in\mathcal{P}([m])} (|\pi| - 1)!.$$

设 $|\pi| = k$ 为块数。umber of blocks. If $S(m,\ k)$ 是将包含 m 个对象的集合划分为 o k 个非空子集 ，那么

$$\sum_{\pi\in\mathcal{P}([m])} (|\pi| - 1)! = \sum_{k=1}^{m} (k - 1)!S(m, k).$$

利用标准上界 $S(m,\ k) \leqslant k^m/k!$ ，我们得到

$$(k - 1)!S(m, k) \leqslant (k - 1)!\frac{k^m}{k!} = k^{m-1},$$

因此

$$\sum_{k=1}^{m} (k - 1)!S(m, k) \leqslant \sum_{k=1}^{m} k^{m-1} \leqslant m^m.$$

最后，利用粗略界限 $m^m \leqslant e^m m!$ 可得

$$\sum_{\pi\in\mathcal{P}([m])} (|\pi| - 1)! \leqslant e^m m!,$$

因此总的来说

$$|\kappa_m(Z)| \leqslant B^m e^m m! = m!(eB)^m.$$

这就证明了（6.23）。

□

引理的证明 *6.14*。 我们使用 H¨older 不等式（相等指数情形）。对于（6.26），我们对 H¨older 不等式使用权重 p_t，对每个 $t \in B$，其中我们对所有 $p_t = s$ 选择 $t \in B$。于是，$\sum_{t\in B} 1/p_t = s \cdot (1/s) = 1$, 因此

$$\mathbb{E}|Z_B| = \mathbb{E}\prod_{t\in B}|Z_t| \leqslant \prod_{t\in B}\left(\mathbb{E}|Z_t|^s\right)^{1/s}.$$

由于 $s \leqslant m$ 和 L^p 范数随 p 单调递增,

$$\left(\mathbb{E}|Z_t|^s\right)^{1/s} \leqslant \left(\mathbb{E}|Z_t|^m\right)^{1/m} \leqslant L$$

对每个 t 均成立。因此

$$\mathbb{E}|Z_B| \leqslant \prod_{t\in B} L = L^s.$$

对于第二个等式，应用柯西–施瓦茨不等式:

$$\mathbb{E}\big[|Z_B|1_{A_n^c}\big] \leqslant \left(\mathbb{E}|Z_B|^2\right)^{1/2}\sqrt{\varepsilon_n}.$$

现在 $|Z_B|^2 = \prod_{t\in B}|Z_t|^2$，因此再次应用指数为 s 的 H¨older 不等式,

$$\mathbb{E}|Z_B|^2 = \mathbb{E}\prod_{t\in B}|Z_t|^2 \leqslant \prod_{t\in B}\left(\mathbb{E}|Z_t|^{2s}\right)^{1/s}.$$

由于 $2s \leqslant 2m$,

$$\left(\mathbb{E}|Z_t|^{2s}\right)^{1/(2s)} \leqslant \left(\mathbb{E}|Z_t|^{2m}\right)^{1/(2m)} \leqslant U,$$

因此

$$\left(\mathbb{E}|Z_t|^{2s}\right)^{1/s} = \left(\left(\mathbb{E}|Z_t|^{2s}\right)^{1/(2s)}\right)^2 \leqslant U^2.$$

Thus

$$\mathbb{E}|Z_B|^2 \leqslant \prod_{t\in B} U^2 = U^{2s},$$

因此

$$\mathbb{E}\big[|Z_B|1_{A_n^c}\big] \leqslant (U^{2s})^{1/2}\sqrt{\varepsilon_n} = U^s\sqrt{\varepsilon_n}.$$

□

7 递减 FP 蕴含 MMSE 上界的证明（定理 2.9的证明）

本节中，我们仅研究超出两个先验分布样本之间重叠范围的随机变量分位数。为此，我们给出一个比定义 2.1更一般的定义。

定义 7.1（分位数函数）。设 Y_n 为一个实值随机变量。对于任意 $D > 0$，我们将分位数 $q(D)$ 定义为 $q(D) := \inf\left\{y \in \mathrm{R} \mid \mathrm{P}(|Y_n| \leqslant y) \geqslant 1 - e^{-D}\right\}$.

我们还要指出以下说明。

注释 7.2. 请注意，定义分位数函数的另一种方式 $q(D)$（这与定义7.1等价）是将 $|Y_n|$的累积分布函数的广义逆函数按时间重新缩放，即 $q(D) = F^{-1}_{|Y_n|}(1 - e^{-D})$（更多详情请参阅，例如 [VdV00，第 21 章。1]）。

7.1 Hermite 背景

我们考虑 R^N 上的概率论学家 *Hermite* 多项式。对于多重指标 $\alpha \in \mathrm{N}^N$ 和向量 $x \in \mathrm{R}^N$，多元 Hermite 多项式 $H_\alpha(x)$ 通过张量积定义：

$$H_\alpha(x) = \prod_{j=1}^{N} h_{\alpha_j}(x_j), \tag{7.1}$$

wh其中 $h_k(z) = (-1)^k e^{z^2/2} \frac{d^k}{dz^k} e^{-z^2/2}$ 表示次数为 k.

这些多项式构成希尔伯特空间 $L^2(\mathrm{R}^N, \gamma)$ 的正交基，其中 γ 是标准高斯测度。内积定义为 $\langle f, g \rangle_{L^2} = \mathbb{E}_Z[f(Z)g(Z)]$，其中 $Z \sim \mathcal{N}(0, I_N)$。正交关系如下：

$$\mathbb{E}_Z[H_\alpha(Z)H_\beta(Z)] = \alpha!\delta_{\alpha,\beta}, \tag{7.2}$$

其中 $\delta_{\alpha,\beta}$ 是克罗内克 delta，且 $\alpha! = \prod_{j=1}^{d} \alpha_j!$。

类似地，对于多重指标 $\alpha \in \mathrm{N}^N$ 和两个向量 $x, y \in \mathrm{R}^N$，滥用符号记法，我们记

$$x^\alpha = \prod_{i=1}^{n} x_i^{\alpha_i} \quad \text{and} \quad (xy)^\alpha = \prod_{i=1}^{n} (x_i y_i)^{\alpha_i}.$$

为推导证明所需的性质，我们将依赖以下基本的平移恒等式，该恒等式由命题 7.3给出，它将平移后的 Hermite 多项式表示为以原点为中心的基底的线性组合。为完整起见，我们将在第 B 节给出证明。

命题 7.3（平移恒等式）。对于 z，$\mu \in \mathrm{R}^N$ 和 $\alpha \in \mathrm{N}^N$：

$$H_\alpha(z + \mu) = \sum_{\gamma \leqslant \alpha} \binom{\alpha}{\gamma} \mu^{\alpha-\gamma} H_\gamma(z). \tag{7.3}$$

利用这一展开，我们证明以下引理。

引理 7.4（高斯均值平移）。设 $Z \sim \mathcal{N}(0, I_N)$ 和 $\mu \in \mathrm{R}^N$。则：

$$\mathbb{E}_Z[H_\alpha(\mu + Z)] = \mu^\alpha. \tag{7.4}$$

证明。将命题 7.3 应用于 $H_\alpha(Z + \mu)$，可得：

$$\mathbb{E}_Z[H_\alpha(Z + \mu)] = \sum_{\gamma \leqslant \alpha} \binom{\alpha}{\gamma} \mu^{\alpha-\gamma} \mathbb{E}_Z[H_\gamma(Z)].$$

根据厄米特多项式的正交性，对于所有 $\gamma \neq 0$，$\mathbb{E}[H_\gamma(Z)] = \langle H_\gamma, H_0 \rangle_{L^2} = 0$。唯一不为零的项对应于 $\gamma = 0$（其中 $H_0 \equiv 1$）。因此：

$$\mathbb{E}_Z[H_\alpha(Z + \mu)] = \binom{\alpha}{0} \mu^\alpha \cdot 1 = \mu^\alpha.$$

□

引理 7.5（平移乘积的期望）。设 $Z \sim \mathcal{N}(0, I_N)$ 和 $\mu \in \mathrm{R}^N$。则：

$$\mathbb{E}_Z\left[H_\alpha(\mu + Z)H_\beta(\mu + Z)\right] = \sum_{r \leqslant \min(\alpha,\beta)} r! \binom{\alpha}{r} \binom{\beta}{r} \mu^{\alpha+\beta-2r}. \tag{7.5}$$

证明。 我们利用平移恒等式（命题 7.3）展开两个多项式：

$$\mathbb{E}_Z\left[H_\alpha(Z + \mu)H_\beta(Z + \mu)\right] = \mathbb{E}_Z\left[\left(\sum_{\gamma \leqslant \alpha} \binom{\alpha}{\gamma} \mu^{\alpha-\gamma} H_\gamma(Z)\right)\left(\sum_{\delta \leqslant \beta} \binom{\beta}{\delta} \mu^{\beta-\delta} H_\delta(Z)\right)\right].$$

利用期望的线性性，我们可以交换求和与期望。由正交关系（7.2）可知， $\gamma \neq \delta$ 对应的交叉项为零。我们只对两个展开式中共有的指标求和 r （$r \leqslant \alpha$ 和 $r \leqslant \beta$）：

$$\begin{aligned}\mathbb{E}_Z\left[H_\alpha(\mu+Z)H_\beta(\mu+Z)\right] &= \sum_{r\leqslant\min(\alpha,\beta)} \binom{\alpha}{r}\mu^{\alpha-r}\binom{\beta}{r}\mu^{\beta-r}\mathbb{E}_Z[H_r(Z)^2] \\ &= \sum_{r\leqslant\min(\alpha,\beta)} \binom{\alpha}{r}\mu^{\alpha-r}\binom{\beta}{r}\mu^{\beta-r}(r!) \\ &= \sum_{r\leqslant\min(\alpha,\beta)} r!\binom{\alpha}{r}\binom{\beta}{r}\mu^{\alpha+\beta-2r}.\end{aligned}$$

□

7.2 关键引理

为了证明低次相关性下界，我们构造并分析一个近乎最优的低次多项式。为此，我们需要几个关键引理，现于此陈述，并将其证明推迟到第 7.4 节。首先，下面给出一个简单的观察：

引理 7 .6（相关性单调不减）。$\mathrm{Corr}_{P_0}^{\leqslant D_n}(h)$ 是关于 $h>0$ 的单调不减函数。

接下来，我们需要给出一个定量误差界，比较指数函数的 D 次泰勒展开与指数函数在零点附近的截断。该引理是构造最优低次多项式的一项基础工具，尤其能解释先验分布重叠的分位数如何出现在分析中。

引理 7.7 （截断指数矩的定量误差界）。设 V 是一个矩存在至 $2D_n+2$ 阶的随机变量，其中 $D_n \in \mathbb{N}$ 是一个次数， $C>0$ 是一个常数， $(q(D'))_{D'>0}$ 是 $|V|$ 在定义 7.1 中给出的分位数。那么，

$$\left|\mathbb{E}[V\exp_{\leqslant D_n}(V)] - \mathbb{E}[Ve^V 1_{|V|\leqslant q(CD_n)}]\right| \leqslant \frac{e^{q(CD_n)}\,\|V\|_{D_n+2}^{D_n+2}}{(D_n+1)!} + e^{-CD_n/2}\sum_{k=0}^{D_n}\frac{\|V\|_{2k+2}^{k+1}}{k!} \tag{7.6}$$

其中 $\|V\|_p = (\mathbb{E}[|V|^p])^{1/p}$ 表示 V 的 L_p-范数。此外，若存在常数 C_1、$C_2>0$，使得 上确界 $\sup_{p\in[2D_n+2]}\|V\|_p \leqslant C_1$ 且 $q(CD_n)\leqslant C_2$，则

$$\left|\mathbb{E}[V\exp_{\leqslant D_n}(V)] - \mathbb{E}[Ve^V 1_{|V|\leqslant q(CD_n)}]\right| = o(e^{-CD_n/2}). \tag{7.7}$$

接下来，我们需要一个引理，用来比较三个同分布随机变量之和的分位数与其中每个随机变量的分位数。

引理 7.8 (分位数支配)。设 X， X'， X'' 是来自先验分布 P_0， $\tilde{q}(D)$ 为和 $S=|\langle X',\ X''\rangle|+|\langle X,\ X'\rangle|+|\langle X,\ X''\rangle|$ 根据定义 7.1得到的分位数函数，且 $q(D)$ 为 $\langle X,\ X'\rangle$ 相应的分位数函数。存在一个普适常数 $C\geqslant 1$，使得对所有 $D\geqslant \ln 3$,

$$q(D)\leqslant \tilde{q}(D)\leqslant Cq(D). \tag{7.8}$$

我们还会用到一个有趣且简单的事实：对于任意函数相对于任意概率测度的期望，都可以用该测度的分位数来表示。

引理 7.9 (变量替换)。设随机变量 $X\geqslant 0$，其分位数函数 q 按定义 7.1定义。对于任意可积函数 g，即 $\mathbb{E}|g(X)|<\infty$，有

$$\mathbb{E}g(X) = \int_0^\infty g(q(t))e^{-t}\mathrm{d}t. \tag{7.9}$$

最后，我们还需要利用随机变量分位数的增长速率来对其进行一般的 L_p 范数控制。

引理 7.10 (矩增长界)。 设 V 为非负随机变量。假设其分位数函数 $q(t)$ 满足上增长条件 $q(t) \leqslant CB_n t^\kappa$ 对所有 $t \geqslant t_0$均成立。则存在常数K，使得对于任意 $p \in \mathbf{N}$，其 L_p 范数均满足：

$$\|V\|_p \leqslant KB_n p^\kappa \tag{7.10}$$

其中常数 K 仅取决于 C 和 κ。

7.3 下界证明

定理 2.9的证明。 根据引理 7.6，相关性$\mathrm{Corr}_{P_0}^{\leqslant D_n}(h)$ 是 $h > 0$中的单调不减函数，并且根据（2.6）以及假设 2.8，

$$\lambda + \frac{d}{dq}\mathcal{F}_{\mathrm{ann},\lambda}\Big|_{q=q(D_n)} = -\frac{d}{dq}\log \mathbb{P}(\langle X, X'\rangle = t)\Big|_{q=q(D_n)} \geqslant \frac{c}{q(D_n)},$$

只需证明

$$\mathrm{Corr}_{P_0}^{\leqslant D_n}\left(\frac{c}{q(D_n)}\right) \geqslant \frac{C}{D_n^\kappa}\sqrt{q(D_n)}. \tag{7.11}$$

为证明这一点，我们采用概率方法，并基于从先验分布中抽取的参考集构造一个多项式估计量，该估计量以高概率满足（7.11）。

令 α 表示 $\mathbf{N}^n$ 中的一个多重指标元素，且 $\lambda = \sqrt{c/q(D_n)}$ 为下文中的元素。我们通过次数为 D_n 的截断厄米展开定义函数$W(Y|X)$：

$$W(Y|X) = \sum_{|\alpha|\leqslant D_n} \frac{1}{\alpha!} X^\alpha H_\alpha(Y). \tag{7.12}$$

令 $M \in \mathbf{N}$ 为一个充分大的数，稍后我们将对其进行选择；$\mathcal{S} = \{X_1, \ldots, X_M\}$ 为从先验分布 P_0 中抽取的一组 M 个独立样本。考虑由下式定义的 D_n 次多项式估计量：

$$p(Y) = \frac{1}{M}\sum_{k=1}^{M} W(Y|\sqrt{\lambda}X_k)\sqrt{\lambda}X_k,$$

给定 X，我们利用命题 7.3 得到

$$\mathbb{E}_Z[H_\alpha(Y)|X] = \mathbb{E}_Z[H_\alpha(\sqrt{\lambda}X + Z)|X] = (\sqrt{\lambda}X)^\alpha.$$

于是，直接代数运算可得：

$$\begin{aligned}
\frac{1}{\sqrt{\lambda}}\mathbb{E}\langle p(Y), X\rangle &= \frac{1}{M}\mathbb{E}_{Z,X}\left[\sum_{k=1}^{M} W(Y|\sqrt{\lambda}X_k)\langle X_k, X\rangle\right] = \frac{1}{M}\sum_{k=1}^{M}\sum_{|\alpha|\leqslant D_n}\frac{1}{\alpha!}(\sqrt{\lambda}X_k)^\alpha \mathbb{E}_{Z,X}[\langle X_k, X\rangle H_\alpha(Y)] \\
&= \frac{1}{M}\sum_{k=1}^{M}\sum_{|\alpha|\leqslant D_n}\frac{1}{\alpha!}\mathbb{E}_X[\langle X_k, X\rangle(\sqrt{\lambda}X_k)^\alpha(\sqrt{\lambda}X)^\alpha] = \frac{1}{M}\sum_{k=1}^{M}\sum_{|\alpha|\leqslant D_n}\frac{1}{\alpha!}\mathbb{E}_X[\langle X_k, X\rangle(\lambda X_k X)^\alpha] \\
&= \frac{1}{M}\sum_{k=1}^{M}\mathbb{E}_X[\langle X_k, X\rangle \exp_{\leqslant D_n}(\lambda\langle X_k, X\rangle)].
\end{aligned}$$

由于 X_k，$k \in [M]$ 是从同一先验分布 P_0 中独立采样的，根据强大数定律，只要 $|\mathbb{E}_{X,\ X'}[\langle X,\ X'\rangle \exp_{\leqslant D_n}(\lambda\langle X,\ X'\rangle)]| < +\infty$，上述结论就以高概率成立，

$$\left|\frac{1}{M}\sum_{k=1}^{M}\mathbb{E}_X[\langle X_k, X\rangle \exp_{\leqslant D_n}(\lambda\langle X_k, X\rangle)] - \mathbb{E}_{X,X'}[\langle X, X'\rangle \exp_{\leqslant D_n}(\lambda\langle X, X'\rangle)]\right| = o_M(1). \tag{7.13}$$

综合以上结果，我们得出：

$$\frac{1}{\sqrt{\lambda}}\mathbb{E}\langle p(Y),X\rangle=\mathbb{E}[\langle X',X\rangle\exp_{\leqslant D_n}(\lambda\langle X',X\rangle)]+o_M(1). \tag{7.14}$$

现在我们调用引理 7.5 给出的恒等式，得到

$$\mathbb{E}_Z[H_\alpha(\sqrt{\lambda}X+Z)H_\beta(\sqrt{\lambda}X+Z)]|X=\sum_{r\leqslant\min(\alpha,\beta)}r!\binom{\alpha}{r}\binom{\beta}{r}(\sqrt{\lambda}X)^{\alpha+\beta-2r}.$$

因此，经过直接代数运算

$$\begin{aligned}
\frac{1}{\lambda}\mathbb{E}[\|p(Y)\|^2]&=\frac{1}{M^2}\sum_{i,j=1}^M\mathbb{E}_{X,Z}[W(Y|\sqrt{\lambda}X_i)W(Y|\sqrt{\lambda}X_j)\langle X_i,X_j\rangle]\\
&=\frac{1}{M^2}\sum_{i,j=1}^M\langle X_i,X_j\rangle\sum_{|\alpha|\leqslant D_n,\,|\beta|\leqslant D_n}\frac{(\sqrt{\lambda}X_i)^\alpha(\sqrt{\lambda}X_j)^\beta}{\alpha!\beta!}\mathbb{E}_{X,Z}[H_\alpha(Y)H_\beta(Y)]\\
&=\frac{1}{M^2}\sum_{i,j=1}^M\langle X_i,X_j\rangle\sum_{\substack{\alpha,\beta,r\\|\alpha|\leqslant D_n,|\beta|\leqslant D_n}}\frac{1}{r!(\alpha-r)!(\beta-r)!}(\sqrt{\lambda}X_i)^\alpha(\sqrt{\lambda}X_j)^\beta\mathbb{E}_X[(\sqrt{\lambda}X)^{\alpha+\beta-2r}]\\
&=\frac{1}{M^2}\sum_{i,j=1}^M\langle X_i,X_j\rangle\sum_{\substack{\alpha,\beta,r\\|\alpha|\leqslant D_n,|\beta|\leqslant D_n}}\frac{1}{r!(\alpha-r)!(\beta-r)!}\mathbb{E}_X[(\lambda X_iX)^{\alpha-r}(\lambda X_jX)^{\beta-r}(\lambda X_iX_j)^r]\\
&=\frac{1}{M^2}\sum_{i,j=1}^M\langle X_i,X_j\rangle\sum_{|r|\leqslant D_n}\frac{1}{r!}\mathbb{E}_X[\exp_{\leqslant D_n-|r|}(\lambda\langle X_i,X\rangle)\exp_{\leqslant D_n-|r|}(\lambda\langle X_j,X\rangle)](\lambda X_iX_j)^r\\
&=\frac{1}{M^2}\sum_{i,j=1}^M\langle X_i,X_j\rangle\sum_{k\leqslant D_n}\frac{1}{k!}\mathbb{E}_X[\exp_{\leqslant D_n-k}(\lambda\langle X_i,X\rangle)\exp_{\leqslant D_n-k}(\lambda\langle X_j,X\rangle)](\lambda\langle X_i,X_j\rangle)^k.
\end{aligned}$$

我们使用以下三个观察：对于给定的 $D\in\mathbb{N}$，$\exp_{\leqslant D}(x)$ 在所有 $x\in\mathbb{R}$ 上单调递增；对于所有 $x\geqslant 0$，$\exp_{\leqslant D}(x)$ 在所有 $D\in\mathbb{N}$ 上单调递增；对于任意 $x,\ y\geqslant 0$，任意 $D_1,\ D_2\in\mathbb{N}$，$\exp_{\leqslant D_1}(x)\exp_{\leqslant D_2}(y)\leqslant\exp_{\leqslant D_1+D_2}(x+y)$。定义

$$h(x,y)=|\langle x,y\rangle|\mathbb{E}_X[\exp_{\leqslant 3D_n}(\lambda(|\langle x,X\rangle|+|\langle y,X\rangle|+|\langle x,y\rangle|))],$$

那么，上述量可以界定如下：

$$\begin{aligned}
&\frac{1}{M^2}\sum_{i,j=1}^M|\langle X_i,X_j\rangle|\sum_{k\leqslant D_n}\frac{1}{k!}\mathbb{E}_X[\exp_{\leqslant D_n-k}(\lambda|\langle X_i,X\rangle|)\exp_{\leqslant D_n-k}(\lambda|\langle X_j,X\rangle|)]|\lambda\langle X_i,X_j\rangle|^k\\
&\leqslant\frac{1}{M^2}\sum_{i,j=1}^M|\langle X_i,X_j\rangle|\mathbb{E}_X[\exp_{\leqslant D_n}(\lambda|\langle X_i,X\rangle|)\exp_{\leqslant D_n}(\lambda|\langle X_j,X\rangle|)]\sum_{k\leqslant D_n}\frac{1}{k!}|\lambda\langle X_i,X_j\rangle|^k\\
&=\frac{1}{M^2}\sum_{i,j=1}^M|\langle X_i,X_j\rangle|\exp_{\leqslant D_n}(|\lambda\langle X_i,X_j\rangle|)\mathbb{E}_X[\exp_{\leqslant D_n}(\lambda|\langle X_i,X\rangle|)\exp_{\leqslant D_n}(\lambda|\langle X_j,X\rangle|)]\\
&\leqslant\frac{1}{M^2}\sum_{i,j=1}^M|\langle X_i,X_j\rangle|\mathbb{E}_X[\exp_{\leqslant 3D_n}(\lambda(|\langle X_i,X\rangle|+|\langle X_j,X\rangle|+|\langle X_i,X_j\rangle|))]=\frac{1}{M^2}\sum_{i,j=1}^M h(X_i,X_j).
\end{aligned}$$

出于同样的原因，由于 $X_k,\ k\in[M]$ 是从相同的先验分布 P_0中独立抽样得到的，应用 U 统计量强大数定律 [VdV00，问题 12.15]，只要 $|\mathbb{E}_{X',\ X''}[h(X',\ X'')]|<+\infty$，

$$\left|\frac{1}{M^2}\sum_{i,j=1}^M h(X_i,X_j)-\mathbb{E}_{X',X''}[h(X',X'')]\right|=o_M(1). \tag{7.15}$$

简而言之，我们证明了以高概率，

$$\frac{1}{\lambda}\mathbb{E}[\|p(Y)\|^2] \leqslant \mathbb{E}_{X,X',X''}[|\langle X', X''\rangle| \exp_{\leqslant 3D_n}\left(\lambda(|\langle X', X''\rangle| + |\langle X, X'\rangle| + |\langle X, X''\rangle|)\right)] + o_M(1).$$

记总和为 $S = |\langle X',\ X''\rangle| + |\langle X,\ X'\rangle| + |\langle X,\ X''\rangle|$，重叠为 $A = \langle X,\ X'\rangle$，则上式可由内积之和的指数函数的 $3D_n$ 次截断控制：

$$|\langle X', X''\rangle| \exp_{\leqslant 3D_n}\left(\lambda(|\langle X', X''\rangle| + |\langle X, X'\rangle| + |\langle X, X''\rangle|)\right) \leqslant S \exp_{\leqslant 3D_n}(\lambda S). \tag{7.16}$$

因此，结合（7.14），存在一个足够大的 M，使得相关性比值以高概率满足以下下界，

$$\mathrm{Corr}_{P_0}^{\leqslant D_n}(\lambda) \geqslant \frac{\mathbb{E}\langle p(Y), X\rangle}{\sqrt{\mathbb{E}[\|p(Y)\|^2]}} \geqslant \frac{\mathbb{E}[A \exp_{\leqslant D_n}(\lambda A)]}{2\sqrt{\mathbb{E}[S \exp_{\leqslant 3D_n}(\lambda S)]}}.$$

我们将引理 7.7 应用于分子 $\mathbb{E}[A\exp_{\leqslant D_n}(\lambda A)]$，因此定义 $\varepsilon_1(D_n) > 0$，使得

$$\left|\mathbb{E}[A \exp_{\leqslant D_n}(\lambda A)] - \mathbb{E}\left[Ae^{\lambda A} 1_{|A|\leqslant q(C_t D_n)}\right]\right| = \varepsilon_1(D_n).$$

类似地，由于 $\lambda = c/q(D_n)$, 定义 $\varepsilon_2(D_n) > 0$ 使得 $|\mathbb{E}[S\exp_{\leqslant 3D_n}(\lambda S)] - \mathbb{E}[Se^{\lambda S}1_{S\leqslant \tilde{q}(C_t D_n)}]| = \varepsilon_2(D_n)$。然后应用引理 7.7 和 7.8，并根据假设 2.8(1)，存在某个常数 $C' > 0$，使得

$$\mathbb{E}[S \exp_{\leqslant 3D_n}(\lambda S)] \leqslant \mathbb{E}[Se^{\lambda S}1_{S\leqslant \tilde{q}(C_t D_n)}] + \varepsilon_2(D_n) \leqslant e^{\lambda \tilde{q}(C_t D_n)}\tilde{q}(C_t D_n) + \varepsilon_2(D_n) \leqslant e^{C'} C' \cdot q(D_n) + \varepsilon_2(D_n).$$

我们将验证 $\varepsilon_1(D_n) = \varepsilon_2(D_n) = o_n(q(D_n))$，使得它们相较于 $q(D_n)$ 在n较大时可以忽略。从 $\varepsilon_1(D_n)$开始，我们采用缩放 $\lambda = c/q(D_n)$。缩放后的变量为 $Z = \lambda A$。根据假设2.8(1)和引理7.10，对于任意 $p \in \mathbb{N}$满足 $p \leqslant 2D_n + 2$,

$$\|Z\|_p = \frac{c\|A\|_p}{q(D_n)} \leqslant \frac{cKB_n p^\kappa}{cB_n D_n^\kappa} \leqslant C_1\left(\frac{p}{D_n}\right)^\kappa \leqslant C_1 3^\kappa.$$

因此，$\sup_{p\in[2D_n+2]}\|Z\|_p \leqslant C_1 3^\kappa$。由于 Z 只是随机变量 A的重新缩放，$|Z|$ 的分位数函数为 $q_{|Z|}(t) = \lambda q(t)$，我们有

$$q_{|Z|}(C_t D_n) = \frac{cq(C_t D_n)}{q(D_n)} \leqslant \frac{cCB_n(C_t D_n)^\kappa}{cB_n D_n^\kappa} = CC_t^\kappa.$$

根据引理7.7，$\varepsilon_1(D_n) = o(e^{-C_t D_n/2})$。根据我们的假设，$D_n = \omega(1)$，$q(D_n) \geqslant cB_n D_n^\kappa = \omega(e^{-CD_n})$。选择截断常数 C_t（在积分范围的定义中）使得 $C_t = 20C > 0$ 足够大，则这验证了 $\varepsilon_1(D_n) = o(q(D_n))$。

同样，引理7.8使得$\tilde{q}(t)$关于$q(t)$的四分位界其中存在一个常数(1) 中给出的一致。按照相同的逻辑，$Z = \lambda S$ 也满足 $\sup_{p\in[2D_n+2]}\|Z\|_p \leqslant C_1$ 且 $q_{|Z|}(C_t D_n) \leqslant C_2$ 对于某些常数$\widetilde{C}_1$，$\widetilde{C}_2 > 0$。由引理7.7，$\varepsilon_2(D_n) = o(e^{-C_t D_n/2}) = o(q(D_n))$。

由于 $|\mathbb{E}[A\exp_{\leqslant D_n}(\lambda A)]| \leqslant |\mathbb{E}[Ae^{\lambda A}1_{|A|\leqslant q(C_t D_n)}]| + \varepsilon_1(D_n) \leqslant q(C_t D_n)e^{\lambda q(C_t D_n)} + \varepsilon_1(D_n) < +\infty$，我们已验证$\mathbb{E}_{X,X'}[\langle X,\ X'\rangle \exp_{\leqslant D_n}(\lambda\langle X,\ X'\rangle)]| < +\infty$。验证$|\mathbb{E}_{X',\ X''}[h(X',\ X'')]| < +\infty$也是一样的。因此，(7.13) 和 (7.15) 均成立。

因此，下界可简化为

$$\mathrm{Corr}_{P_0}^{\leqslant D_n}(\lambda) \geqslant c\frac{\mathbb{E}\left[Ae^{\lambda A}1_{|A|\leqslant q(D_n)}\right]}{\sqrt{q(D_n)}} - o(1).$$

由于 $xe^x \geqslant x + \frac{1}{2}x^2$ 对所有 $x \in [-1,\ 1]$ 都成立，代入 $x = \lambda A$ 并截断于事件 $|x| \leqslant \lambda q(D_n) \leqslant 1$ 上，我们得到

$$\mathbb{E}\left[Ae^{\lambda A}1_{|A|\leqslant q(D_n)}\right] \geqslant \mathbb{E}[A1_{|A|\leqslant q(D_n)}] + \frac{1}{2}\lambda\mathbb{E}[A^2 1_{|A|\leqslant q(D_n)}] \geqslant \mathbb{E}[A] - |\mathbb{E}[A1_{|A|\geqslant q(D_n)}]| + \frac{1}{2}\lambda\mathbb{E}[A^2 1_{|A|\leqslant q(D_n)}].$$

代入 $g(|A|)=|A|^2 1_{|A|\leqslant q(D_n)}$ 引理7.9，我们得到

$$\mathbb{E}[A^2 1_{|A|\leqslant q(D_n)}]=\int_0^{\infty} q(t)^2 1_{q(t)\leqslant q(D_n)} e^{-t}\mathrm{d}t=\int_0^{D_n} q(t)^2 e^{-t}\mathrm{d}t.$$

根据假设 2.8(1)， $c_1 B_n t^{\kappa}\leqslant q(t)$ 对于所有 $t\leqslant D_n$，存在一个新的绝对常数 $c'>0$，使得

$$\begin{aligned}\int_0^{D_n} q(t)^2 e^{-t}\mathrm{d}t &\geqslant c_1 B_n^2\int_0^{D_n} t^{2\kappa}e^{-t}\mathrm{d}t = c_1 B_n^2\left(\Gamma(2\kappa+1)-\int_{D_n}^{\infty} t^{2\kappa}e^{-t}\mathrm{d}t\right)\\ &\geqslant c_1 B_n^2\left(\Gamma(2\kappa+1)-CD_n^{2\kappa+1}e^{-D_n}\right)\geqslant c' B_n^2.\end{aligned}$$

根据我们对 A 的定义， $\mathbb{E}[A]=\mathbb{E}_{X,\ X'\sim P_0,\ X\perp X'}[\langle X,\ X'\rangle]=\|\mathbb{E}X\|^2\geqslant 0$，并利用引理 7.10，$|\mathbb{E}[A1_{|A|\geqslant q(D_n)}]|\leqslant\sqrt{\mathbb{E}[A^2]\mathbb{P}(|A|\geqslant q(D_n))}\leqslant e^{-D_n/2}\mathbb{E}[A^2]\leqslant Ce^{-D_n/2}B_n^2$。综上，我们得到

$$\mathbb{E}\left[Ae^{\lambda A}1_{|A|\leqslant q(D_n)}\right]\geqslant c' B_n^2(\lambda-C'e^{-D_n/2}).$$

由于 $\lambda=c/q(D_n)\geqslant cB_n^{-1}D_n^{-\kappa}=\omega(e^{-D_n/2})$, 最终，存在一个常数 $C>0$ 使得

$$\mathrm{Corr}_{P_0}^{\leqslant D_n}(\lambda)\geqslant\frac{c'}{2}\cdot\frac{B_n D_n^{-\kappa}}{\sqrt{q(D_n)}}\geqslant CD_n^{-2\kappa}\sqrt{q(D_n)}.$$

□

7.4 关键引理的证明

引理的证明 7.6 $0<h_1<h_2$ $\mathrm{MMSE}_{P_0}^{\leqslant D_n}(h_1^2)\leqslant \mathrm{MMSE}_{P_0}^{\leqslant D_n}(h_2^2)$ $Y_1=h_1X+$。。我们将证明对任意，相关性 相关性。记 $Z\ Y_2=h_2X+Z$ $W\perp\!\!\!\perp Y_2\ W\sim N(0\ I_N)$ $X\perp\!\!\!\perp Z\sim N(0\ I_N)$ 作，。由于，，那么存在一个随机向量，，，

$$Y_1\overset{d}{=}\frac{h_1}{h_2}Y_2+\sqrt{1-\frac{h_1^2}{h_2^2}}W\,.$$

根据 [SW22，断言 A.2], 加入噪声只会增大最小均方误差，因此可得

$$\mathrm{MMSE}_{P_0}^{\leqslant D}(Y_1)=\mathrm{MMSE}_{P_0}^{\leqslant D}\left(\frac{h_1}{h_2}Y_2+\sqrt{1-\frac{h_1^2}{h_2^2}}W\right)\geqslant\mathrm{MMSE}_{P_0}^{\leqslant D}\left(\frac{h_1}{h_2}Y_2\right)=\mathrm{MMSE}_{P_0}^{\leqslant D}(Y_2).$$

因此，经过简单的代数运算可得

$$(\mathrm{Corr}_{P_0}^{\leqslant D})^2(h_1^2)=\mathbb{E}_{X\sim P_0}[\|X\|_2^2]-\mathrm{MMSE}_{P_0}^{\leqslant D}(Y_1)\leqslant\mathbb{E}_{X\sim P_0}[\|X\|_2^2]-\mathrm{MMSE}_{P_0}^{\leqslant D}(Y_2)=(\mathrm{Corr}_{P_0}^{\leqslant D})^2(h_2^2).$$

□

引理 7.7 的证明。 令 Δ 表示所关注的量:

$$\Delta=\left|\mathbb{E}[V\exp_{\leqslant D_n}(V)]-\mathbb{E}[Ve^V 1_{|V|\leqslant q(CD_n)}]\right|.$$

我们将多项式项的期望分解为在事件 $E=\{|V|\leqslant q(CD_n)\}$ 及其补集 $E^c=\{|V|>q(CD_n)\}$上的期望:

$$\mathbb{E}[V\exp_{\leqslant D_n}(V)]=\mathbb{E}[V\exp_{\leqslant D_n}(V)1_E]+\mathbb{E}[V\exp_{\leqslant D_n}(V)1_{E^c}].$$

将其代入 Δ 并应用三角不等式:

$$\Delta\leqslant\underbrace{\mathbb{E}\left[|V|\cdot\left|\exp_{\leqslant D_n}(V)-e^V\right|1_E\right]}_{T_1}+\underbrace{\mathbb{E}\left[|V|\cdot\left|\exp_{\leqslant D_n}(V)\right|1_{E^c}\right]}_{T_2}.$$

对 T_1进行界定：根据带有拉格朗日余项的泰勒定理，对于任意 v，存在介于 0 和 v之间的 ξ，使得 $e^v - \exp_{\leqslant D_n}(v) = \frac{e^\xi}{(D_n+1)!}\, v^{D_n+1}$。在事件 ~~$E =$~~ $\{|V| \leqslant q(CD_n)\}$上，我们有 $|V| \leqslant q(CD_n)$。由于 $q(CD_n) > 0, \xi \leqslant \max(0,\ V) \leqslant q(CD_n)$，这意味着

$$\left|\exp_{\leqslant D_n}(V) - e^V\right| 1_E \leqslant \frac{e^{q(CD_n)}}{(D_n+1)!}|V|^{D_n+1} 1_E.$$

将其代入 T_1，并放宽示性函数 $1_E \leqslant 1$:

$$T_1 \leqslant \mathbb{E}\left[|V|\frac{e^{q(CD_n)}}{(D_n+1)!}|V|^{D_n+1}\right] = \frac{e^{q(CD_n)}}{(D_n+1)!}\|V\|_{D_n+2}^{D_n+2}.$$

对 T_2 进行界定：展开多项式并使用三角不等式:

$$T_2 = \mathbb{E}\left[\left|\sum_{k=0}^{D_n}\frac{V^{k+1}}{k!}\right| 1_{|V|>q(CD_n)}\right] \leqslant \sum_{k=0}^{D_n}\frac{1}{k!}\mathbb{E}\left[|V|^{k+1}1_{|V|>q(CD_n)}\right].$$

我们对期望项应用柯西–施瓦茨不等式:

$$\mathbb{E}\left[|V|^{k+1}1_{|V|>q(CD_n)}\right] \leqslant \sqrt{\mathbb{E}[|V|^{2k+2}]}\sqrt{\mathbb{E}[1^2_{|V|>q(CD_n)}]} = \|V\|_{2k+2}^{k+1}\sqrt{\mathbb{P}(|V|>q(CD_n))}.$$

利用尾部假设 $\mathbf{P}(|V| > q(CD_n)) \leqslant e^{-CD_n}$，可得 $\sqrt{\mathbf{P}(|V| > q(CD_n))} \leqslant e^{-CD_n/2}$。因此:

$$T_2 \leqslant e^{-CD_n/2}\sum_{k=0}^{D_n}\frac{\|V\|_{2k+2}^{k+1}}{k!}.$$

此外，若 $\sup_{p\in[2D_n+2]}\|V\|_p \leqslant C_1$ 且 $q(CD_n) \leqslant C_2$，则利用界 $(D_n+1)! \geqslant (D_n/e)^{D_n}$, 存在一个足够大的常数 $C_3 > 0$,

$$T_1 \leqslant \frac{e^{C_2}\|V\|_{D_n+2}^{D_n+2}}{(D_n+1)!} \leqslant e^{C_2}C_1^2\left(\frac{eC_1}{D_n}\right)^{D_n} = o\left(e^{-C_3 D_n}\right);$$
$$T_2 \leqslant e^{-CD_n/2}\sum_{k=0}^{D_n}\frac{\|V\|_{2k+2}^{k+1}}{k!} \leqslant e^{-CD_n/2}\sum_{k=0}^{D_n}\frac{C_1^{k+1}}{k!} \leqslant C_1 e^{C_1}e^{-CD_n/2} = o(e^{-CD_n/2}).$$

因此在这种情况下，$\Delta = o(e^{-CD_n/2})$。 □

引理7.8的证明。由于 $S = |\langle X',\ X''\rangle| + |\langle X,\ X'\rangle| + |\langle X,\ X''\rangle|$。令 V_1，V_2，V_3 为这些同分布的项，它们与 $|\langle X,\ X'\rangle|$具有相同的分布。利用并集界:

$$\mathbb{P}(S > 3q(D)) \leqslant \sum_{k=1}^{3}\mathbb{P}(V_k > q(D)) = 3e^{-D} = e^{-(D-\ln 3)}. \tag{7.17}$$

由于 $q(t)$ 是递增的，$q(D-\ln 3) \leqslant q(D)$。因此，我们可以选择 C，使事件 $\{S \leqslant Cq(D)\}$ 以至少 $1-e^{-D}$的概率成立，这意味着 $\tilde{q}(D) \leqslant Cq(D)$。□

引理的证明 7.9。根据定义 7.1给出的分位数函数定义，我们知道 $q(t) = F_X^{-1}(1-e^{-t})$ 对所有 $t \geqslant 0$。接着使用 [VdV00，引理 21。1]，$F_X^{-1}(U) \sim X$对于 $U \sim$均匀分布$[0,\ 1]$。因此，对于任意可积函数 g，我们有

$$\mathbb{E}g(X) = \mathbb{E}g(F_X^{-1}(U)) = \int_0^1 g(F_X^{-1}(u))\mathrm{d}u = \int_0^\infty g(F_X^{-1}(1-e^{-t}))e^{-t}\mathrm{d}t = \int_0^\infty g(q(t))e^{-t}\mathrm{d}t.$$

□

引理的证明 7.10。 我们利用尾积分表示来表示 p 阶矩，并应用由（7.9）给出的变量替换公式，积分变为：

$$\mathbb{E}[V^p] = \int_0^\infty (q(t))^p e^{-t} \mathrm{d}t.$$

由于 $q(t) \leqslant CB_n t^\kappa$ ，并通过一个常数 M 处理较小的 t ：

$$\mathbb{E}[V^p] \leqslant \int_0^\infty (CB_n t^\kappa)^p e^{-t} \mathrm{d}t = (CB_n)^p \int_0^\infty t^{\kappa p} e^{-t} \mathrm{d}t = (CB_n)^p \Gamma(\kappa p + 1). \tag{7.18}$$

取 t 第 p次根，并使用 [Ver18第 26]页给出的伽马函数上界：

$$(\Gamma(\kappa p + 1))^{1/p} \leqslant (C(\kappa p/e)^{\kappa p})^{1/p} = C^{1/p}(\kappa p/e)^\kappa. \tag{7.19}$$

因此，

$$\|V\|_p = (\mathbb{E}[V^p])^{1/p} \leqslant CB_n(\kappa/e)^\kappa p^\kappa. \tag{7.20}$$

令 $K = C(\kappa/e)^\kappa,$ ，即可得到所需的上界 $\|V\|_p \leqslant KB_n p^\kappa$。 □

8 具有正累积量的两大类 GAM

本节将介绍两个关键引理，帮助我们验证假设2.4中的条目 1，即感兴趣的 GAM 先验分布具有低阶非负累积量。事实上，我们将证明两大类 GAM 均满足这一条件，其中也包括应用中我们感兴趣的模型作为特例。我们在此以一定的普遍性陈述这些结果，它们本身也具有独立的研究价值。

更具体地说，本节考虑的所有 GAM 均满足以下条件。

固定整数 $t \geqslant 1$。假设我们有来自某个分布 $P_1 \in \mathcal{P}(\mathbb{R})$ 的 $v_1, \ldots, v_n \sim P_1$ 个独立同分布抽样，该分布具有有限阶矩；并且 $Z_1, \ldots, Z_t$ 是一些独立同分布随机变量，对所有 $j \in [t]$，它们都关于 0 对称，或满足 $Z_j \geqslant 0, j \in [t]$ 几乎必然成立。此外，对于所有 $i \in [n]$ 和 $j \in [t]$， Z_j 与 v_i 相互独立。于是我们假设，对所有 $1 \leqslant i \leqslant N$，

$$X_i = Z_{\alpha(i)} \cdot f_i(v_1, \ldots, v_n) \tag{8.1}$$

其中 f_i： $\mathbb{R}^n \to \mathbb{R}$ 是关于 v_1、……、 v_n的单项式，即 $f(v_1$，……， $v_n) = v_1^{k_1} \ldots v_n^{k_n}$其中 k_1、……、 k_n 是某些正整数，且 $a(i) \in [t]$ 对所有 $i \in [n]$均成立。当然，所有张量 PCA 模型 [MR14] 都属于这一框架。

8.1 两类 GAM

8.1.1 P_1 具有非负且超乘性矩

首先考虑关于先验分布 P_1的以下假设。

假设 8.1. 设对于某个 $c > 0$ ，先验分布 P_1 满足以下假设。

- （“非负矩”）对于所有 $k \in \mathbb{N}$ 和 $v \sim P_1$ ，均有 $\mathbb{E}_{P_1}[v^k] \geqslant 0$。
- （“超乘性矩”）对于 $\rho = n^{-c}$，对于任意 $k,$ ，均有 $\mathbb{E}_{P_1}[v^k]\mathbb{E}_{P_1}[v^t] \leqslant \rho \cdot \mathbb{E}_{P_1}[v^{k+t}]$ $t \in \mathbb{N}$.

对于所有满足假设 P_1 8.1 的 GAM，我们可以证明它们具有低阶非负累积量。

结果如下，其证明推迟至第 8.2 节。

引理 8.2。 假设我们观测到一个形式如下的 *GAM* (8.1)，其先验分布为 P_1 ，满足假设*8.1*，其中 $c > 0$。那么，对于任意 $\alpha \in \mathbb{N}^n$ 使得 $1 \leqslant |\alpha| \leqslant$ 对数$_2(1/\rho) - 1,$ ，其中 $\rho = n^{-c}$，我们有

$$\kappa_\alpha(X_1, X_2, \ldots, X_n) \geqslant 0,$$

Now，有趣的是，假设 8.1 对于先验分布 P_1 被许多标准先验分布所满足。

引理 8.3。 设 $\rho = n^{-c}$ 其中 $c > 0$。那么考虑 $B_n \sim$ 伯努利分布(ρ) 以及 T 任意一个与 B_n 独立的实值随机变量，且其关于 0 对称分布，或几乎处处非负。那么，若 P_1 是 $T \times B_n$的分布， P_1 则满足假设 *8.1* 对于 $c > 0$.

特别地， $B_n \sim$伯努利分布(ρ) 和 $R_n \sim Rad(\rho)$ 满足假设 *8.1*。

证明。 设 $v \sim P_1$。在引理的设定下，我们可以写成 $v = T_n := B_n T$，其中 $B_n \sim$ 伯努利分布(ρ) 与 T 独立。

首先，很容易看出，在关于 T 的任一条件下，对于每个 $m \in \mathbf{N}$，都有 $\mathbb{E}[T^m] \geqslant 0$。此外， $B_n \geqslant 0$ 几乎处处成立，因此对于每个 $m \in \mathbf{N}$， $\mathbb{E}[v^m] \geqslant 0$ ，且假设的第一个条件得到满足。

因此，我们只需证明假设的第二个条件。任取 k， $t \in \mathbf{N}$ 和任意 $m \in \mathbf{N}$。由于 $B_n \in \{0, 1\}$ 我们有 $B_n^m = B_n$，因此由独立性可得，

$$\mathbb{E}_{P_1}[v^m] = \mathbb{E}[(B_n T)^m] = \mathbb{E}[B_n^m]\,\mathbb{E}[T^m] = \mathbb{E}[B_n]\mathbb{E}[T^m] = \rho\,\mathbb{E}[T^m]. \tag{8.2}$$

利用这一点可得

$$\mathbb{E}_{P_1}[v^k]\mathbb{E}_{P_1}[v^t] = \rho^2\ \mathbb{E}[T^k]\mathbb{E}[T^t], \qquad \mathbb{E}_{P_1}[v^{k+t}] = \rho\ \mathbb{E}[T^{k+t}].$$

因此，所需的不等式

$$\mathbb{E}_{P_1}[v^k]\mathbb{E}_{P_1}[v^t] \leqslant \rho\,\mathbb{E}_{P_1}[v^{k+t}]$$

等价于

$$\mathbb{E}[T^k]\mathbb{E}[T^t] \leqslant \mathbb{E}[T^{k+t}]. \tag{8.3}$$

若 $T \geqslant 0$ 几乎处处成立，则由 H¨older 不等式（指数为 $\frac{k+t}{k}$ 和 $\frac{k+t}{t}$），

$$\mathbb{E}[T^k] = \mathbb{E}\left[(T^{k+t})^{\frac{k}{k+t}}\right] \leqslant \mathbb{E}[T^{k+t}]^{\frac{k}{k+t}}, \qquad \mathbb{E}[T^t] = \mathbb{E}\left[(T^{k+t})^{\frac{t}{k+t}}\right] \leqslant \mathbb{E}[T^{k+t}]^{\frac{t}{k+t}}.$$

两边相乘得到（8.3）。

如果 T的分布关于 0 对称，那么只要 k、t中至少有一个是奇数，就有 $\mathbb{E}[T^k]\mathbb{E}[T^t] = 0$。此外，若 $\mathbb{E}[T^{k+t}] = 0$为奇数，则 $k + t$；而若 $\mathbb{E}[T^{k+t}] \geqslant 0$为偶数，则 $k + t$，因为此时 $T^{k+t} \geqslant 0$几乎处处成立。因此（8.3）成立。若 k 和 t 均为偶数，令 $Y := T^2 \geqslant 0$。于是

$$\mathbb{E}[T^k]\mathbb{E}[T^t] = \mathbb{E}[Y^{k/2}]\,\mathbb{E}[Y^{t/2}] \leqslant \mathbb{E}[Y^{(k+t)/2}] = \mathbb{E}[T^{k+t}].$$

因此，在所有情况下（8.3）均成立，从而证明了 P_1的条件。

特别地，取 $T = 1$ 几乎处处成立，可得 $v = B_n$；再取 T 为一个与 B_n 独立的 Rad$(1/2)$ 随机变量，可得 $v = B_n T$，其分布为 Rad(ρ)，因此 B_n 和 $R_n \sim \mathrm{Rad}(\rho)$ 满足假设 8.1。 □

8.1.2 P_1 具有非负累积量

现在我们转向第二类 GAM。

首先，对于分布 $\mu \in \mathcal{P}(\mathbf{R})$ 以及任意 $r \in \mathbf{N}$，我们将 r阶累积量记为 μ 的

$$\kappa_r(X) := \kappa(\underbrace{X, X, \ldots, X}_{r-\text{times}}),$$

其中 $X \sim \mu$。现在，第二个假设要求 P_1 的所有 r阶累积量均为非负。

假设 8。 4（非负累积量）。设随机变量 $X \sim P_1$ 对任意 $r \in \mathbf{N}$ ，

$$\kappa_r(X) \geqslant 0.$$

注意，假设 8.4 被多种标准分布满足，例如高斯分布 $\mathcal{N}(\mu, \sigma^2)$其中 $\mu \geqslant 0$，泊松分布 Pois(λ) 对任意 $\lambda > 0$，以及指数分布 Exp(λ) 对任意 $\lambda > 0$；详见例如 [McC87，第 2] 章。

现在，假设我们有如下形式的 GAM（8.1），其中 $Z_j = 1$ 对所有 $j \in [t]$均成立，即

$$X_i = f_i(v_1, \dots, v_n) \tag{8.4}$$

其中 f_i： $\mathbf{R}^n \to \mathbf{R}$ 是关于独立同分布随机变量的单项式 v_1，…，v_n。若分布P_1满足 v_i，$i \in [n]$ 则满足假设 8.4 时，以下结论成立。注意，该引理直接推出低阶累积量非负性质，同时还给出了一个有技术用途的第二个条件。

引理 8.5。对于具有如下形式的任意 *GAM* (8.4)，若其分布 $P_1 \in \mathcal{P}(\mathbf{R})$ 满足假设 *8.4*，则以下结论成立。对于所有 $\alpha \in \mathbf{N}^N$,

$$\kappa_\alpha(X_1, \dots, X_N) \geqslant 0,$$

以及 $\tilde{\kappa}_\alpha(X_1、\dots、X_N) \geqslant \kappa_\alpha^2(X_1、\dots、X_N)$。

该引理的证明推迟至第 8.3 节

8.2 引理 8.2 的证明

对于 本节剩余部分提醒读者，为方便起见，我们已定义 ，

$$\kappa_\alpha(X_1, \mathbf{X}) = \kappa_\alpha(X_1, \underbrace{X_1, \dots, X_1}_{\alpha_1-\text{times}}, \dots, \underbrace{X_N, \dots, X_N}_{\alpha_N-\text{times}}).$$

继续证明之前，先给出一个定义。

定义 8.6（依赖多重图）。对于我们关注的 GAM 类，注意到对于任意 $\kappa_\alpha(X_1、\dots...、X_N)$，多重指标 $\alpha \in \mathbf{R}^N$ 自然对应一个“依赖”多重图，其顶点为 $\{1、\dots...、n\}$；其中对于任意 $i \in [N]$，若 $\alpha_i \neq 0$，则该多重图在满足 $v_j \in \operatorname{supp}(f_i)$ 的顶点 $j \in [n]$ 之间有一个团。

现在陈述一个对证明有用的推论，它可由命题 6.2立即得到。 .

推论 8.7。对于任意 $\alpha \in \mathbf{R}^N$，若其依赖多重图不连通，则 $\kappa_\alpha(X_1，\dots...，X_N) = 0$。

因此，我们可以只考虑依赖多重图连通的多重指标 $\alpha \in \mathbf{N}^n$。

现在我们继续证明。

引理的证明 *8.2*。为方便起见，我们假设 $\lambda = 1$ 因为所有条件相对于λ都是齐次的。

另外请注意，为方便起见，我们将证明，对于某个固定的 $\kappa_\alpha(X_i，\mathbf{X}) \geqslant 0$ 以及所有基数为 $1 \leqslant |\alpha| \leqslant \log_2(1/\rho) - 1.$ 的多重指标 α，都有 $i \in [n]$。结合以下事实：对于所有$i \in [n]$，$\kappa(X_i) \geqslant 0$ 成立；由于 X_i 对所有 $i \in [n]$ 都是独立随机变量 v_i，$i \in [n]$ 的幂的乘积，且这些随机变量具有非负矩，因此对于所有满足 $1 \leqslant |\alpha| \leqslant \log_2(1/\rho) - 1$ 的 α，都有 $\kappa_\alpha(X_1，\dots...，X_n) \geqslant 0$，这正是所需结果。

在证明余下部分时，不失一般性，我们固定 $i = 1$，并证明 $\kappa_\alpha(X_1，\mathbf{X}) \geqslant 0$，从而得出所需结果。

考虑任意一个形式为（8.1）的 GAM，其先验分布为 P_1，满足假设 8.1，其中$c > 0$。我们首先通过对 $|\alpha|$ 进行归纳法来证明以下结论：

$$\frac{1}{2}\mathbb{E}[X_1 X^\alpha] \leqslant \kappa_\alpha \leqslant \mathbb{E}[X_1 X^\alpha]. \tag{8.5}$$

注意，对于 $\alpha = \mathbf{0}$，$\kappa_0 = \mathbb{E}[X_1] \geqslant 0$，根据假设 8.1。

现在假设 归纳假设对任意满足 β 且 $|\beta| < |\alpha|$ 的情形都成立。利用命题 6.5，我们知道

对于所有多重指标中 $\alpha \in \mathbf{N}^N$ 累积量 κ_α 满足以下递归公式，

$$\kappa_\alpha = \mathbb{E}[X_1 X^\alpha] - \sum_{0 \leqslant \beta < \alpha} \kappa_\beta \binom{\alpha}{\beta} \mathbb{E}[X^{\alpha-\beta}]. \tag{8.6}$$

然而，归纳假设意味着，对于所有 P_1 满足假设 8.1 的情形，都有 $\kappa_\beta \geqslant 0$ 对于所有 $\beta \prec \alpha$，因为 $X_1 X^\alpha$ 可以表示为独立 v_i的幂的乘积，而 v_i的矩均为非负数。这意味着第二项非正，因此我们得到 $\kappa_\alpha \leqslant \mathbb{E}[X_1 X^\alpha]$。为了完成归纳步骤，我们需要证明 $\kappa_\alpha \geqslant \frac{1}{2}\mathbb{E}[X_1 X^\alpha]$。对于下界，我们利用假设 8.1 ，按照以下方式对 κ_α 的递归公式中的第二项取上界：首先

$$\sum_{0 \leqslant \beta < \alpha} \kappa_\beta \binom{\alpha}{\beta} \mathbb{E}[X^{\alpha-\beta}] \leqslant \sum_{0 \leqslant \beta < \alpha} \binom{\alpha}{\beta} \mathbb{E}[X_1 X^\beta] \mathbb{E}[X^{\alpha-\beta}]$$

其中我们再次使用了归纳假设。我们要证明，对于每个多重指标 $0 \leqslant \beta \prec \alpha$，都有

$$\mathbb{E}[X_1 X^\beta] \mathbb{E}[X^{\alpha-\beta}] \leqslant \mathbb{E}[X_1 X^\alpha] \rho. \tag{8.7}$$

不过请注意， X_i 具有8.1 的形式，而 Z_j， $j \in [t]$ 与 v_i， $i \in [n]$相互独立。因此，为了证明我们所需的不等式，只需证明对所有这样的 β ，都有

$$\mathbb{E}[Z_i^k] \mathbb{E}[Z_i^t] \leqslant \mathbb{E}[Z_i^{k+t}] \quad \text{and} \quad \mathbb{E}[v_j^{\beta_j}] \mathbb{E}[v_j^{\alpha_j - \beta_j}] \leqslant \rho \mathbb{E}[v_j^{a_j}]$$

对所有 $i \in [t]$ ，且至少有一个 $j \in [n]$ （我们有一组独立同分布随机变量的乘积，因此可以分别处理每个坐标）。对于第一个不等式，请注意，这可由以下事实推出： Z_j， $j \in [t]$几乎处处对称或非负——我们也在引理8.3中证明了这一点。对于第二个不等式，我们如下处理。由于 α 是一个连通多重图，因此对任意满足 $0 < \beta < \alpha$ 的 $\kappa_\beta > 0$ ，都有 β 也是连通的。此外，

$$|V(\beta)| + |V(\alpha - \beta)| \geqslant |V(\alpha)| + 1. \tag{8.8}$$

确实，如果情况并非如此，那么 $V(\alpha - \beta)$ 和 $V(\alpha)$ 的顶点将构成 α 的两个互不连通的连通分支，这将导致矛盾。方程（8.8）意味着总是存在一个坐标 $j \in [n]$ 使得 $j \in (\{1\} \cup \operatorname{supp}(\beta)) \cap \operatorname{supp}(\alpha)$。对于这个坐标，表达式

$$\mathbb{E}[X_1 X^\beta] \mathbb{E}[X^{\alpha-\beta}]$$

包含这一项 $\mathbb{E}[v_j^{\beta_j}]\mathbb{E}[v_j^{\alpha_j-\beta_j}]$ $\beta_j > 0$ 8 1 $\mathbb{E}[v_j^{\beta_j}]\mathbb{E}[v_j^{\alpha_j-\beta_j}] \leqslant$，其中 。利用假设 。，我们可以给出上界，$\mathbb{E}[v^{\alpha_j}]\rho$ 8 7 并且（ 。）的证明即告完成。利用这个方程，我们得到：

$$\begin{aligned}
\sum_{0 \leqslant \beta < \alpha} \binom{\alpha}{\beta} \mathbb{E}[X_1 X^\beta] \mathbb{E}[X^{\alpha-\beta}] &\leqslant \sum_{0 \leqslant \beta < \alpha} \binom{\alpha}{\beta} \mathbb{E}[X_1 X^\alpha] \rho \\
&\leqslant 2^{|\alpha|} \mathbb{E}[X_1 X^\alpha] \rho \\
&\leqslant \frac{1}{2} \mathbb{E}[X_1 X^\alpha].
\end{aligned}$$

这里我们利用了 $|\alpha| \leqslant \log_2(1/\rho) - 1$ 这一事实，得到最后一个不等式。该上界意味着$\kappa_\alpha \geqslant \frac{1}{2}\mathbb{E}[X_1 X^\alpha]$ ，归纳法的证明至此完成。

现在利用8.5，我们知道，对于任意先验分布 P_1 满足假设8.1 对于所有 α 使得 $|\alpha| \leqslant C$ 对数 n且对于某个普适常数 $C := C(\rho)$ 该先验分布的累积量 $\kappa(X_1, \mathbf{X})$ 满足 $\kappa(X_1, \mathbf{X}) \geqslant 0$ 证明完毕。

□

8.3 引理8.5的证明

引理*8.5*的证明。 我们将再次使用定理6.17。

给定任意形式为（8.1）的 GAM，我们希望将任意形式的累积量分解为 $\kappa_\alpha(f_1, \ldots\ldots, f_N)$，其中 $\alpha \in \mathbf{N}^N$。由于 f_i对所有 $i \in [N]$均成立，等于 $f_i(v_1, \ldots\ldots, v_n) = v_1^{k_1^i} \ldots v_n^{k_n^i}$，其中 $k_1^i, \ldots\ldots, k_n^i$ 为某些正整数，我们应用定理 6.17 于其中的

$$\sum_{j=1}^{N} \alpha_j \Big(\sum_{i=1}^{n} k_i^j \Big)$$

出现在该累积量中的随机变量（注意，在这一组随机变量中，许多 v_i 会重复出现多次，但在这一论证中，我们仍将它们视为不同的随机变量）。进一步说明如下：对于 $i=1$、 $f_1(v_1$、……、$v_n)$ 是 $S_1=\sum_{i=1}^n k_i^1$ 个随机变量的乘积。将这些变量重新标记为 R_1^1、……、$R_{S_1}^1$，并对所有 f_i， $i\in[n]$重复相同的过程后，我们可以将所关注的累积量改写为：

$$\kappa_\alpha(f_1,\dots,f_n)=\kappa_\alpha(R_1^1\cdot\dots\cdot R_{S_1}^1,\dots,R_1^N\cdot\dots\cdot R_{S_n}^N)$$

于是，定理6.17 给出了如下形式的分解

$$\kappa_\alpha(f_1,\dots,f_N)=\sum_{\pi'\in G}\prod_{S'\in\pi'}\kappa(S')$$

其中值得注意的是，右侧出现的所有累积量都具有以下形式

$$\kappa(v_{i_1},\dots v_{i_t}),\qquad \text{for some}\quad i_1,\dots,i_t\in[n],t\geqslant 1,$$

其中 $i_1,\dots,i_t$ 不一定互不相同。我们指出，唯一非零的项具有以下形式$\kappa(v_i,\dots,v_i)$ 其中某个 $i\in[n]$。事实上，如果存在某一项对应于一个部分 S'，其中至少包含两个$v_i,v_j\in S'$，且对应于两个不同的整数 $i,j\in[n]$，那么 $\kappa(S')=0$，因为 v_i,v_j 相互独立。使用我们上文引入的符号说明可知，这意味着求和中只会出现先验分布在 $r-$上的第 v_i 阶累积量，其中 $r\in\mathbf{N}$。假设 8.4 保证这些项均非负，因此对所有 $\alpha\in\mathbf{N}^N$

$$\kappa_\alpha(f_1,\dots,f_N)\geqslant 0.$$

最后，利用引理 6.18 我们还可得出 $\tilde{\kappa}_\alpha(X_1,\dots X_N)\geqslant\kappa_\alpha^2(X_1,\dots X_N)$ 证明至此完成。□

9 修正贝塞尔函数背景

在本附录中，设 $d\geqslant 1$ 为整数，且 G， $H\in\mathbf{R}^d$ 为相互独立的 $N(0,\ I_d)$ 向量。定义

$$W:=\langle G,H\rangle=\sum_{i=1}^d G_iH_i,$$

并令 f_d 表示 的密度。关于修正贝塞尔函数的背景知识，参见 [YC17、AS65]。对于 $\nu\in\mathbf{R}$ ，第一类修正贝塞尔函数定义为

$$I_\nu(x):=\sum_{m=0}^\infty\frac{1}{m!\,\Gamma(m+\nu+1)}\left(\frac{x}{2}\right)^{2m+\nu},\qquad x>0.$$

特别地，对于 $\nu=0$，我们有积分表示

$$I_0(x)=\frac{1}{\pi}\int_0^\pi\exp(x\cos\theta)d\theta.$$

对于 $\nu=0$，积分表示与级数定义的等价性可直接由指数函数的泰勒展开得到。展开 $\exp(x\cos\theta)$，并利用一致收敛交换求和与积分，可得 $\frac{1}{\pi}\sum_{k=0}^\infty\frac{x^k}{k!}\int_0^\pi\cos^k\theta d\theta$。奇数 k 时，$\cos^k\theta$ 在 $[0,\ \pi]$ 上的积分由对称性可知为零。对于偶次幂 $k=2m$， 利用标准三角积分可得 $\int_0^\pi\cos^{2m}\theta d\theta=\pi\frac{(2m)!}{2^{2m}(m!)^2}$。代入将此结果代入求和式后， π和 $(2m)!$ 项相互抵消，恰好得到 $\sum_{m=0}^\infty\frac{1}{(m!)^2}\left(\frac{x}{2}\right)^{2m}$。由于 $\Gamma(m+1)=m!$，这正好还原了 $I_0(x)$的级数定义。

9.1 第二类修正贝塞尔函数

第二类修正贝塞尔函数随后对 $\nu \notin \mathbb{Z}$ 定义如下：

$$K_\nu(x) := \frac{\pi}{2\sin(\pi\nu)}\big(I_{-\nu}(x) - I_\nu(x)\big), \qquad x > 0,$$

整数阶情形则由连续性得到，参见 [YC17，(1.1)]。

它还具有如下积分表示

$$K_\nu(x) = \int_0^\infty e^{-x\cosh u}\cosh(\nu u)\,du, \qquad x > 0,$$

参见 [YC17，(1.2)] 以及 [AS65，第 9]章。特别地，$K_\nu(x) > 0$ 对所有 $x > 0$，且 $K_{-\nu}(x) = K_\nu(x)$。我们将使用以下标准恒等式。

引理 9.1（K_ν 的递推关系和导数公式）。对每个 $x > 0$ 以及每个实数 ν，

$$K_{\nu+1}(x) = K_{\nu-1}(x) + \frac{2\nu}{x}K_\nu(x), \tag{9.1}$$

$$\frac{d}{dx}\big(x^\nu K_\nu(x)\big) = -x^\nu K_{\nu-1}(x), \tag{9.2}$$

$$K'_\nu(x) = -K_{\nu-1}(x) - \frac{\nu}{x}K_\nu(x) = -K_{\nu+1}(x) + \frac{\nu}{x}K_\nu(x), \tag{9.3}$$

$$\frac{d}{dx}\log K_\nu(x) = -\frac{\nu}{x} - \frac{K_{\nu-1}(x)}{K_\nu(x)} = \frac{\nu}{x} - \frac{K_{\nu+1}(x)}{K_\nu(x)}. \tag{9.4}$$

等价地，

$$\frac{d}{dx}\log\big(x^\nu K_\nu(x)\big) = -\frac{K_{\nu-1}(x)}{K_\nu(x)}. \tag{9.5}$$

证明。递推关系（9.1）和导数恒等式（9.2）是标准结果，参见 [AS65，(9.6.26)]。展开（9.2）可得（9.3）中的第一个公式，而第二个公式可由第一个公式与（9.1）结合得到。除以 $K_\nu(x)$得到（9.4），而（9.5）只需对（9.2）取对数导数即可得到。□

我们使用的小参数和大参数渐近式如下。

引理 9.2（基本渐近式）。固定 $\nu > 0$。当 $x \downarrow 0$，

$$K_\nu(x) = \frac{(1+o(1))}{2}\Gamma(\nu)\Big(\frac{x}{2}\Big)^{-\nu}, \tag{9.6}$$

而 $K_0(x) = -(1+o(1))$对数 x. 当 $x \to \infty$，

$$K_\nu(x) = (1+o(1))\sqrt{\frac{\pi}{2x}}e^{-x}\left(1 + \frac{4\nu^2-1}{8x} + O_\nu(x^{-2})\right). \tag{9.7}$$

因此，对任意 $\nu_1, \nu_2 = O(1)$，

$$\frac{K_{\nu_1}(x)}{K_{\nu_2}(x)} \to 1 \qquad as\ x \to \infty.$$

证明。这些都是标准结果，参见 [YC17，(1.3)–(1.5)] 和 [AS65，第 9]。比值极限可由共同的首项 $\sqrt{\pi/(2x)}e^{-x}$立即得到。□

为便于后续使用，我们还记录两个比值估计。

引理 9.3（比值不等式）。对于所有 $\nu \geqslant 0$ 和 $x > 0$，

$$\frac{\nu + \sqrt{x^2+\nu^2}}{x} < \frac{K_{\nu+1}(x)}{K_\nu(x)} < \frac{\nu + \frac{1}{2} + \sqrt{x^2 + (\nu+\frac{1}{2})^2}}{x}. \tag{9.8}$$

证明。 不等式（9.8）恰好是 [YC17，(1.10)]。 □

推论 9.4（比值界的大阶推论）。令 $\nu \to \infty$ 并令 $x = x_\nu > 0$ 满足 $x = o(\nu)$。那么

$$\frac{K_{\nu-1}(x)}{K_\nu(x)} = \frac{x}{2\nu}\big(1 + o(1)\big). \tag{9.9}$$

证明。 将（9.8）中的 ν 替换为 $\nu - 1$:

$$\frac{\nu - 1 + \sqrt{x^2 + (\nu-1)^2}}{x} < \frac{K_\nu(x)}{K_{\nu-1}(x)} < \frac{\nu - \frac{1}{2} + \sqrt{x^2 + (\nu - \frac{1}{2})^2}}{x}.$$

由于 $x = o(\nu)$,

$$\nu - 1 + \sqrt{x^2 + (\nu-1)^2} = 2\nu + o(\nu), \qquad \nu - \frac{1}{2} + \sqrt{x^2 + (\nu - \frac{1}{2})^2} = 2\nu + o(\nu).$$

因此

$$\frac{K_\nu(x)}{K_{\nu-1}(x)} = \frac{2\nu}{x}\big(1 + o(1)\big),$$

取倒数即得（9.9）。 □

9.2 高斯内积的密度

$W = \langle G,\ H\rangle$ 的分布可以用 K_ν 显式表示。

引理 9.5 （高斯内积的显式密度）。 令 $W = \langle G、\ H\rangle = \sum_{i=1}^d G_i H_i$.。则 $W = \frac{X-Y}{2}$，其中 X、Y是独立的 χ^2_d 随机变量。此外， W的密度为

$$f_d(x) = \frac{1}{\sqrt{\pi}\,\Gamma(d/2)\,2^{d/2-1/2}}\,|x|^{\,d/2-1/2}\,K_{\,d/2-1/2}(|x|), \qquad x \in \mathbb{R}\backslash\{0\}. \tag{9.10}$$

对于 $d \geqslant 2$，公式可连续延拓至 $x = 0$。对于 $d = 1$,

$$f_1(x) = \frac{1}{\pi}K_0(|x|),$$

且当 $x \to 0$ 时，$f_1(x) = -\frac{(1+o(1))}{\pi}$ 对数 $|x|$。

该引理的证明推迟至第 10.4节。

10 应用的证明

在证明定理 3.1、3.2 和3.3 之前，我们先陈述一个辅助引理，其证明推迟到第 10.4 节。

引理 10.1 (拉德马赫和的界)。令 $S_n = \sum_{i=1}^n \varepsilon_i$ 满足 $\varepsilon_i \overset{iid}{\sim} \mathrm{Rad}(\pm 1)$ 和 $\phi(x) = (2\pi)^{-1/2}e^{-x^2/2}$。对于整数 $s \equiv n$ (模 2)，以下结论成立。

令 $T = T_n \geqslant 1$ 满足 $T^3 = o(\sqrt{n})$。那么，对于满足 $|s| \leqslant \sqrt{n}\,T$的整数 $s \equiv n$ (模 2)，以下结论一致成立，

$$\mathbb{P}(S_n = s) = \frac{2}{\sqrt{n}}\,\phi\Big(\frac{s}{\sqrt{n}}\Big)\,\big(1 + o(1)\big). \tag{10.1}$$

此外，存在绝对常数 c、$C > 0$ 和 $n_\star \geqslant 2$，使得对所有 $n \geqslant n_\star$ 以及满足$s \equiv n$ (模 2) 且满足 $0 \leqslant s \leqslant n/2$，的所有整数，均有

$$\mathbb{P}(S_n = s) \ \geqslant\ \frac{c}{\sqrt{n}}\exp\Big(-\frac{s^2}{2n} - C\frac{s^3}{n^2}\Big). \tag{10.2}$$

10.1 稀疏张量主成分分析的证明

我们先介绍一些符号说明。设 $v,\ v'\in\mathbb{R}^n$ 以及 $k=n^{\beta+o(1)},\ \beta\in(0,\ 1)$ 的条目服从独立同分布的 Rad 分布(k/n)，即每个条目以概率 $k/n,-1$ 取 1，以概率 k/n 取 -1，否则取 0。定义 $p:=(k/n)^2$。那么，若 $X_1=v_1v_1'$，……，$X_n=v_nv_n'$，这些随机变量独立同分布，且 $\mathbb{P}(X_i=1)=\mathbb{P}(X_i=-1)=p/2,\mathbb{P}(X_i=0)=1-p.$。令 $S:=\sum_{i=1}^n X_i$，以及 $T:=\sum_{i=1}^n 1\{|X_i|=1\}.$。我们记

$$\mu:=\mathbb{E}[T]=np=\frac{k^2}{n}. \tag{10.3}$$

另请注意

$$\mathrm{Var}(X_i)=\mathbb{E}[X_i^2]=p,\qquad \mathrm{Var}(S)=n\mathrm{Var}(X_i)=np=\mu,\qquad \sigma=\sigma_S=\frac{k}{\sqrt{n}}.$$

根据以上所有内容，两个从先验分布中独立同分布抽取的样本的内积 $\langle X,\ X'\rangle=\langle v^{\otimes r},\ v'^{\otimes r}\rangle$ 具有形式 S^r。

在证明本节主定理之前，我们先陈述一些将要用到的引理。对于这些模型中的每一个，我们都需要两个关键要素：确定式（2.6）中导数在阶数 $q(D_n)$ 处的阶数，然后确定分位数 $q(D_n)$的主导阶。在稀疏张量主成分分析中，我们有两个稀疏度区间 $k=\omega(\sqrt{n})$ 和 $k=o(\sqrt{n})$ ，并将分别处理。

为了分析中等稀疏度区间内的算法阈值（其中 $\beta\in(1/2,1),$ ），我们首先需要仔细控制重叠分位数 $q(D)$ 的尺度。下面的引理为该分位数函数给出了紧确的（相差常数因子）上界和下界，这些界随 D 变化。

引理 10.2. 取任意常数 $c_1>0$、整数 $r\geqslant 2$ 和稀疏度 $k=n^{\beta+o(1)}$，其中 $\beta\in(1/2,1)$。那么存在绝对常数 $0<c<C<\infty$ 和 $\mu_0\geqslant 1$（仅依赖于 c_1），使得只要μ（定义见 (10.3)）满足 $\mu\geqslant\mu_0$，且对所有满足 D 的 c_1 对数 $\mu\leqslant D\leqslant\mu$，我们都有

$$\left(c\,\sigma\sqrt{D}\right)^r\leqslant q(D)\leqslant\left(C\,\sigma\sqrt{D}\right)^r.$$

引理 10.2 的证明推迟至第10.1.2节。

基于上述分位数界，我们现在刻画重叠的对数概率的离散导数的衰减率。

引理 10.3. 取 $k=n^{\beta+o(1)}$，其中 $\beta\in(1/2,1)$，且 $D_n\leqslant C$对数 n，其中 $C>0$ 是任意正数常量。那么，对于 $a_n(s)=(s+2)^r-s^r$、 $s\in 2\mathbb{Z}$ 在 $s=q(D_n)$ 处的取值，以下结论成立：

$$-\Delta_{a_n}\log\mathbb{P}(\langle X,X'\rangle=q(D_n))=\Theta\left(\frac{n^{\frac{r}{2}}}{k^r}D_n^{2-r}\right)=\Theta\left(\lambda_{\mathrm{ALG}}\cdot D_n^{2-r}\right),$$

引理 10.3 的证明推迟至第10.1.3节。

接下来，我们将关注高度稀疏区间，其特征为 $\beta\in(0,\ 1/2)$。在这种情况下，下面的引理确定了分位数函数 $q(D_n)$ 的渐近阶。

引理 10.4。固定 $\beta\in(0,1/2)$ 并令 $k=n^{\beta+o(1)}$，于是 $\mu:=k^2/n=n^{-(1-2a)}\to 0$。假设 $D_n=n^{o(1)}$ 且$D_n\geqslant c_0$ 对数 n，其中某个固定的 $c_0>0$。那么存在常数 $c,C>0$（仅依赖于 a 和 c_0），使得对所有充分大的 n，

$$c\frac{D_n}{\log n}\leqslant q(D_n)\leqslant C\frac{D_n}{\log n}.$$

引理 10.4 的证明推迟至第10.1.4节。

本节最后一个引理估计了小 k时对数概率的离散导数，从而完成高度稀疏区间所需的技术性估计。

引理 10.5. 固定一个整数 $r\geqslant 2$, $\beta\in(0,1/2)$，并令 $k=n^{\beta+o(1)}$ 和 $a_n(s):=(s+2)^r-s^r$，$s\in\mathbb{Z}_{\geqslant 0}$。那么，对于任意 c_0 对数 $n\leqslant D_n=n^{o(1)}$，存在某个固定的 $c_0>0$，且 $s_n:=q_n(D_n)$，结论成立

$$-\Delta_{a_n}\log\mathbb{P}(\langle X,X'\rangle=s_n^r)=\Theta\left(\frac{(\log n)^r}{D_n^{r-1}}\right).$$

引理 10.5的证明推迟到第 10.1.5 节。

10.1.1 定理3.2 的证明

我们通过直接应用定理 3.2 来证明定理 2.7。为此，我们首先证明稀疏拉德马赫先验分布满足假设 2.10。具体而言，我们验证这四个条件。

我们从 $k = n^{\beta+o(1)}$, $\beta \in (1/2, 1)$ 的情形开始。

条件 1: 由于该模型的形式为（8.1）（对 $Z_j \equiv 1$ 对所有 $j \in [t]$），且我们指定的先验分布 P_0 满足假设 8.1，利用引理 8.2，我们知道，对于所有多重指标 $\alpha \in \mathbf{N}^n$ 满足 $|\alpha| \leqslant D_n$对于 $D_n = \lfloor$对数$_2(n/k) - 1\rfloor$，有 $\kappa_\alpha(X_i) \geqslant 0$。

条件 2: 由于 X_i 均被 1 所界定，因此该条件显然成立。

条件 3: 该条件显然成立。对于 $D'_n = D_n$对数$^2 n$ 并利用引理 10.2 对于 $k = n^{\beta+o(1)}$, $\beta \in (1/2, 1)$,

$$q(D'_n) \geqslant (D'_n)^{r/2}(k/\sqrt{n})^r = \omega(\max\{\log^{-D_n} n, n^{-C}\}),$$

对任意常数 $C > 0$。

条件 4: 根据引理 10.3 可知，在 $q(D_n)$ 处的导数的阶为 $n^{\frac{r}{2}} D_n^{2-r}/k^r$ 对于 $k = \omega(\sqrt{n})$。因此，假设 4 等价于以下条件:

$$\frac{D_n^{2-r} n^{\frac{r}{2}}/k^r}{A_n} \leqslant \frac{1}{q(\log^3 n)\log n}.$$

根据引理 10.2 可知，对于

$$D'_n = \log^3 n, \tag{10.4}$$

满足 $\omega(\log \mu) = D'_n = o(\mu)$, 的条件下，有

$$q(D'_n) \leqslant C_1 \frac{k^r \log^{3r/2} n}{n^{\frac{r}{2}}}.$$

将此不等式代入上式可得，对于 $A_n = C_1 D_n^{2-r}\log^{(3r+2)/2} n$,, 假设 4 成立，其中 $C_1 > 0$ 为正数常量。

综合以上所有结果以及（2.6），我们已证明，对于序列 $A_n = C_1 D_n^{2-r}$对数$^{(3r+2)/2} n$，有

$$\mathrm{Corr}_{P_0}^{\leqslant D_n}\left(\frac{1}{A_n}\cdot\left(\lambda + \Delta_{a_n}\mathcal{F}_{\mathrm{ann},\lambda}(q(D_n))\right)\right)^2 \leqslant q(D'_n)$$

其中 D'_n 定义于（10.4）中定义，且 a_n 表示引理中定义的序列10.3。现在注意到 at

$$\mathrm{MMSE}_X^{\mathrm{trivial}} = \mathbb{E}\|X\|^2 - (\mathbb{E}\|X\|)^2 = k^r.$$

接下来，结合引理 10.3 和（2.6），可知对于所有 $k = n^\beta, \beta \in (1/2, 1)$, $\lambda + \Delta_{a_n}\mathrm{F}_{\mathrm{ann},\lambda}(q(D_n)) = \Theta\left(\frac{n^{\frac{r}{2}}}{k^r} D_n^{2-r}\right)$。利用这一点并取 $\lambda = \frac{n^{\frac{r}{2}}}{k^r} D_n^{2-r}/$对数 n，可得对于此 λ, $\Delta_{a_n}\mathrm{F}_{\mathrm{ann},\lambda}(q(D_n)) \geqslant 0$。因此，利用定理 2.7，相关性 $(\lambda/A_n)^2 \leqslant 2q(D'_n)$,，其中 $A_n = C_1$对数$^3 n$。等价地，利用引理 10.2 和 6.8，并代入 $D_n = \Theta($对数 $n)$

$$\mathrm{MMSE}_X^{\leqslant D_n}\left(\frac{\lambda}{\log^3 n}\right) \geqslant k^r - C_1\frac{k^r \log^r n}{n^{\frac{r}{2}}}.$$

最后，由于在此区域内，$\lambda_{\mathrm{ALG}} = \widetilde{\Theta}(n^{r/2}/k^r)$:

$$\mathrm{MMSE}_X^{\leqslant D_n}\left(\frac{\lambda_{\mathrm{ALG}}}{\log^{(3r+4)/2} n}\right) \geqslant k^r - C_1\frac{k^r \log^r n}{n^{\frac{r}{2}}}.$$

由于 $k = o(n)$ 我们已经证明了稠密情形下 $k = n^{\beta+o(1)}$, $\beta \in (1/2, 1)$ 以及所有$r \geqslant 2$所需的不等式。

我们知道，对 $k = n^{\beta+o(1)}$, $\beta \in (0, 1/2)$ 证明该结果。

条件 1： 同样，由于该模型具有（8.1）的形式（对于 $Z_j \equiv 1$ 所有 $j \in [t]$），且我们指定的先验分布 P_0 满足 8.1，利用引理 8.2 可知，对于所有满足以下条件的多重指标 $\alpha \in \mathbb{N}^n$，当 $|\alpha| \leqslant D_n$ 时，对于$D_n = \lfloor$对数$_2(n/k) - 1\rfloor$，有 $\kappa_\alpha(X_i) \geqslant 0$。

条件 2：这是因为 X_i 均以 1 为界。
条件 3： 根据引理 10.5 对于 $k = n^{\beta+o(1)}$， $\beta \in (0,\ 1/2)$,

$$q(D_n') \geqslant \Theta(1) = \omega(\max\{\log^{-D_n} n, n^{-C}\}),$$

对于任意常数 $C > 0$。
条件 4 根据引理10.5 我们知道，在 $q(D_n)$ 处的导数的阶为 $\Theta(1)$。因此，假设 4 可化简为:

$$\Theta\left(\frac{1}{A_n}\right) \leqslant \frac{1}{q(\log^3 n)\log n},$$

对于某个正数常量 $C_1 > 0$， $A_n = C_1\log^3 n$ 成立。

综上所述，结合（2.6），我们已证明，对于序列 $A_n = C_1 D_n^{2-r}$对数$^{(3r+2)/2}\, n$有

$$\mathrm{Corr}_{P_0}^{\leqslant D_n}\left(\frac{1}{A_n}\cdot\left(\lambda + \Delta_{a_n}\mathcal{F}_{\mathrm{ann},\lambda}(q(D_n))\right)\right)^2 \leqslant q(D_n')$$

其中 a_n 表示引理 10.5中定义的序列。

接下来，利用引理 10.5和（2.6），可得对所有 $k = n^\beta$， $\beta \in (0,\ 1/2)$, $\lambda + \Delta_{a_n}\mathcal{F}_{\mathrm{ann},\ \lambda}(q(D_n)) =$Θ(对数 n)。利用这一点并取 $\lambda = 1$，可得对于这个 λ， $\Delta_{a_n}\mathcal{F}_{\mathrm{ann},\ \lambda}(q(D_n)) \geqslant 0$。最后，由于在此区间内， $\lambda_{\mathrm{ALG}} = \Theta(1)$ 应用定理 2.7可得:

$$\mathrm{MMSE}_X^{\leqslant D_n}\left(\frac{\lambda_{\mathrm{ALG}}}{\log^3 n}\right) \geqslant k^r - C_1\log^2 n.$$

由于 $k = \omega($对数$^2\, n)$，我们已经证明了稀疏区间 $k = n^{\beta+o(1)}$， $\beta \in (0,\ 1/2)$ 以及所有 $r \geqslant 2$下所需的不等式。

10.1.2 引理 10.2的证明

上界 B ound：由于 $|X_i| \leqslant 1$、 $\mathbb{E}X_i = 0$ 且 $\mathrm{Var}(S) = \mu$，伯恩斯坦不等式给出，对所有 $t \geqslant 0$,

$$\mathbb{P}(|S| \geqslant t) \leqslant 2\exp\Big(-\frac{t^2}{2(\mu + t/3)}\Big).$$

取 $t = C(\sqrt{\mu D} + D)$，对于足够大的绝对常数 C，有 $\mathbf{P}(|S| \geqslant t) \leqslant e^{-D}$。由于 $D \leqslant \mu$，那么 $\sqrt{\mu D} + D \leqslant 2\sqrt{\mu D}$，所以

$$q(D) \leqslant (2C\sqrt{\mu D})^r = (2C\,\sigma\sqrt{D})^r.$$

下界： 设 $A_t = \sum_{j=1}^t \varepsilon_j$ 满足 $\varepsilon_j \overset{iid}{\sim} \mathrm{Rad}(\pm 1)$。固定 $c_1 > 0$，如陈述中所示。由引理 10.1可知，存在常数 $c' > 0$ 和 $t_0 \geqslant 2$ （仅依赖于 c_1），使得对所有 $t \geqslant t_0$ 以及所有 D 满足 c_1 对数 $t \leqslant D \leqslant t$ 的，均有

$$\mathbb{P}\big(|A_t| \geqslant c'\sqrt{tD}\big) \geqslant e^{-D}. \tag{10.5}$$

记 $X_i = \varepsilon_i Z_i$，其中 $\varepsilon_i \sim \mathrm{Rad}(\pm 1)$ 和 $Z_i := \mathbf{1}_{\{|X_i|=1\}} \sim \mathrm{Ber}(\rho)$ 相互独立。则 $T = \sum_i Z_i \sim$二项分布 $(n,\ \rho)$ 满足 $\mathbb{E}T = \mu$，并且在给定 $T = t$ 的条件下，有 $S \overset{d}{=} A_t$。由切尔诺夫不等式,

$$\mathbb{P}(T \leqslant \mu/2) \leqslant e^{-\mu/8} \quad \text{and} \quad \mathbb{P}(T \geqslant 2\mu) \leqslant e^{-\mu/3},$$

因此，当 μ 足够大时，$\mathbf{P}(\mu/2 \leqslant T \leqslant 2\mu) \geqslant 3/4$。在这个事件上，我们将其记为 E， $D \leqslant \mu \leqslant 2T$，因此 $D/2 \leqslant T$并且 $e^{-D/2} \geqslant e^{-D}$。此外， $T \leqslant 2\mu$ 蕴含 $\log\, T \leqslant$对数$(2\mu) \leqslant 2$对数 μ当 μ足够大时，因此 $D/2 \geqslant (c_1/4)$对数 T。因此，引理 10.1 （常数为 $c_1/4$）适用于 $t = T$和 $D/2$。因此，我们可以将 (10.5) 与 $t = T$结合使用，得到

$$\mathbb{P}\Big(|S| \geqslant c'\sqrt{TD}\ \Big|\ T, E\Big) \geqslant \mathbb{P}\Big(|A_t| \geqslant c'\sqrt{TD}\Big) \geqslant e^{-D},$$

这里我们再次利用了：在给定 $T=t$ 的条件下， S 与 A_t 具有相同的分布。利用 $T \geqslant \mu/2$ 得到 $\sqrt{TD} \geqslant \sqrt{\mu D/2},$ ，因此

$$\mathbb{P}\big(|S| \geqslant c\sqrt{\mu D}\big) \;\geqslant\; \mathbb{P}\Big(2\mu \geqslant T \geqslant \frac{\mu}{2}\Big)e^{-D} \;\geqslant\; \frac{3}{4}e^{-D}$$

在调整 c （并将常数吸收到 μ_0 中）之后。因此， $|S|$ 的 $\frac{3}{4}e^{-D}$ 上分位数至少为 $c\sqrt{\mu D}$。提升到 r 次幂可得 $q(D+3/4) \geqslant (c\sqrt{\mu D})^r = (c\sigma\sqrt{D})^r$。等价地，调整一些常数后，对于 $C = c/2$ ，有 $q(D) \geqslant (c\sqrt{\mu(D-3/4)})^r = (c\sigma\sqrt{D-3/4})^r \geqslant (C\sigma\sqrt{D})^r,$ ，因为 $D=\omega(1)$。结合步骤1，证明完成。

10.1.3 引理 10.3 的证明

我们记 $S = \langle v,\ v'\rangle$，因此由两次独立同分布抽样 $X,\ X' \sim P$得到的量 $\langle X,\ X'\rangle$ 为 S^r。利用这一点，量

$$-\Delta_{a_n} \log \mathbb{P}(\langle X, X'\rangle = q(D_n))$$

等于，将 $S^r = \langle X,\ X'\rangle$ 和 $a_n(s) = (s+2)^r - s^r$ 代入，其中整数 $s = q(D_n),\ D_n \leqslant C$ 满足对数 n，且 $C > 0,$

$$\Lambda_r(s) := \frac{\log \mathbb{P}(S = s+2) - \log \mathbb{P}(S = s)}{(s+2)^r - s^r},$$

下面先介绍一些符号说明和辅助引理。

如前所述，我们将计算 $\Lambda_r(s)$ 在 $s = q(D_n)$ 处的值，或者根据引理10.2 ，这等价于计算$s = \Theta(D_n\sigma_S)$处的值。将 s 中的首项常数 D_n 吸收进去后，我们只需计算 $s = D_n\sigma_S$处的值。任取一个确定性序列 $D_n \geqslant 1$ 并定义

$$s_n := 2\left\lfloor \frac{D_n\sigma_S}{2} \right\rceil \quad \text{so that} \quad |s_n - D_n\sigma_S| \leqslant 1 \ \text{ and } \ s_n \equiv 0 \pmod 2.$$

设 $\varepsilon_1, \dots, \varepsilon_t$为独立同分布的 Rademacher 随机变量，$\mathrm{P}(\varepsilon_i = \pm 1) = 1/2,$并定义

$$A_t := \sum_{i=1}^{t} \varepsilon_i.$$

注意，除非 $s \equiv t$ 模 2，否则 $\mathrm{P}(A_t = s) = 0$ 。

我们先计算给定简单随机游走 A_t时步长为 2 的比率。稍后，一旦证明相关的 t 值集中在 μ附近，这将作为展开 S 的主项。

引理 10.6 （对 A_t的展开）。 对满足 $s \equiv t \pmod 2$ 且 $s \leqslant t-2$的整数 $t \geqslant 1$ 和 s ，一致地有，

$$\log \frac{\mathbb{P}(A_t = s+2)}{\mathbb{P}(A_t = s)} = -\frac{2s+2}{t} + O\left(\frac{(s+1)^2}{t^2}\right).$$

证明。如果 $s \equiv t \bmod 2$ 对于 $j := (t+s)/2 \in \{0, \dots, t\},$成立，则有

$$\mathbb{P}(A_t = s) = 2^{-t}\binom{t}{j}.$$

否则 $\mathrm{P}(A_t = s) = 0$。事实上，如果 $B_t := \#\{i :\ \varepsilon_i = +1\}$ ~ 服从二项分布$(t,\ 1/2)$。那么 $A_t = 2B_t - t$ ，且事件 $\{A_t = s\}$ 等价于 $\{B_t = (t+s)/2\}$。

现在，对于满足 $s \equiv t \bmod 2$ 且符合上述条件 $j = (t+s)/2$ 的 $s \in \mathbb{Z}$，有 $s \leqslant t-2$。那么 $(t+s+2)/2 = j+1$，并且

$$\frac{\mathbb{P}(A_t = s+2)}{\mathbb{P}(A_t = s)} = \frac{\binom{t}{j+1}}{\binom{t}{j}} = \frac{t-j}{j+1} = \frac{t - \frac{t+s}{2}}{\frac{t+s}{2} + 1} = \frac{t-s}{t+s+2}.$$

取对数后，

$$\log \frac{\mathbb{P}(A_t = s+2)}{\mathbb{P}(A_t = s)} = \log\left(\frac{t-s}{t+s+2}\right) = \log(1-u), \qquad u := \frac{2s+2}{t+s+2}.$$

若 $s \leqslant t-2$ 则 $0<u \leqslant 1/2$。对于 $u \in [0,\ 1/2]$ 我们有 $|\text{对数}(1-u)+u| \leqslant 2u^2$，因此对数$(1-u) = -u + O(u^2)$。此外

$$u = \frac{2s+2}{t} \cdot \frac{1}{1+\frac{s+2}{t}} = \frac{2s+2}{t} + O\left(\frac{(s+1)^2}{t^2}\right), \qquad u^2 = O\left(\frac{(s+1)^2}{t^2}\right).$$

合并后即可得到所述结论。 □

下一个引理将 S 的分布改写为随机游走 A_t 的分布的混合，其中混合变量是随机支撑重叠 T。这一分解是关键的化简，使我们能够将关于 A_t 的估计转化为关于 S 的估计。

引理 10.7 （二项混合表示）。 我们有 $T = \sum_{i=1}^{n} 1\{|X_i| = 1\} \sim$ 二项分布(n, q)，其中 $q = (k/n)^2$ 且 $\mathbb{E}[T] = \mu = k^2/n$。此外，给定 $T = t$,

$$S \mid (T = t) \overset{d}{=} A_t.$$

因此，对于任意 $s \in \mathbb{Z}$,

$$\mathbb{P}(S = s) = \sum_{t=0}^{n} \mathbb{P}(T = t)\mathbb{P}(A_t = s).$$

证明。 每个 X_i 以概率 q 非零，因此 $T \sim$ 二项分布$(n,\ q)$。给定支撑集的大小（因而给定 $T = t$）后，各符号独立同分布且对称，所以其和是长度为 t 的拉ademacher 随机游走 A_t。对 T 求平均即可得到该公式。□

固定 s，定义权重

$$w_t(s) := \mathbb{P}(T = t)\mathbb{P}(A_t = s), \qquad \text{for all} \quad t = 0, 1, \ldots, n.$$

那么 $\mathrm{P}(S = s) = \sum_t w_t(s)$，并注意到，除非 $t \equiv s \mod 2$，否则 $w_t(s) = 0$ 会自动成立。接下来，假设对某个固定值 $s \in \mathbb{R}$，有 $\mathrm{P}(S = s) > 0$，并定义

$$r_t(s) := \begin{cases} \dfrac{\mathbb{P}(A_t = s+2)}{\mathbb{P}(A_t = s)} & \text{if } \mathbb{P}(A_t = s) > 0, \\ 0 & \text{if } \mathbb{P}(A_t = s) = 0. \end{cases} \tag{10.6}$$

利用混合表示，我们现在将比率 $\mathrm{P}(S = s+2)/\mathrm{P}(S = s)$ 表示为相应条件比率的加权平均，其中 A_t。因此，问题归结为理解哪些 t 取值承载了大部分概率质量，以及 $r_t(s)$ 在这些取值之间变化有多大。

引理 10.8（步长为 2 的比率作为加权平均）。假设$\mathrm{P}(S = s) > 0$ 对某个固定值 $s \in \mathbb{R}$成立。那么，

$$\frac{\mathbb{P}(S = s+2)}{\mathbb{P}(S = s)} = \frac{\sum_{t=0}^{n} w_t(s) r_t(s)}{\sum_{t=0}^{n} w_t(s)}.$$

地，若对所有 $t \geqslant 0$ 都有 $u_t(s) :\, = w_t(s)/\sum_u w_u(s)$，（因此 $u_t(s) \geqslant 0$ 且 $\sum_t u_t(s) = 1$），那么等价

$$\frac{\mathbb{P}(S = s+2)}{\mathbb{P}(S = s)} = \sum_{t=0}^{n} u_t(s) r_t(s), \quad \textit{for all} \quad t = 0, 1, \ldots, n.$$

证明。由引理10.7,

$$\mathbb{P}(S = s+2) = \sum_t \mathbb{P}(T = t)\mathbb{P}(A_t = s+2).$$

如果 $\mathrm{P}(A_t = s) = 0$，那么同样有 $\mathrm{P}(A_t = s+2) = 0$，因此恒等式 $\mathrm{P}(A_t = s+2) = r_t(s)\mathbb{P}(A_t = s)$ 对所有t都成立。因此

$$\mathbb{P}(S = s+2) = \sum_t \mathbb{P}(T = t) r_t(s) \mathbb{P}(A_t = s) = \sum_t w_t(s) r_t(s).$$

除以 $\mathrm{P}(S = s) = \sum_t w_t(s)$。

□

为了有效使用加权平均表示，我们需要知道，当 t 保持在 μ 周围的典型窗口内时，条件比率 $r_t(s)$不会发生太大变化。下面的引理恰好给出了这一稳定性估计。

引理 10.9。 假设 $k = n^{\beta+o(1)}$, $\beta \in (1/2, 1)$ ，特别是 $\mu = k^2/n = \omega(1)$。令 $s \leqslant D_n\sqrt{\mu}$ 并定义

$$W := C'\sqrt{\mu}\log\mu, \qquad \mathcal{I} := \{t \in \{0, \dots, n\} : \ |t-\mu| \leqslant W\},$$

其中 $C' > 0$ 是一个固定常数。那么对于所有满足 $t, t' \geqslant s+2$的 $t, t' \in I$,

$$|r_t(s) - r_{t'}(s)| = O\left(\frac{D_n \log\mu}{\mu}\right),$$

其中 $r_t(s)$, $s \in \mathbf{R}$ 定义见 (10.6)。

证明。 对于实数 $t > s$，令 $f(t) := (t-s)/(t+s+2)$。那么

$$f'(t) = \frac{2s+2}{(t+s+2)^2}.$$

在 $\mathcal{I}$ 上，由于 $D_n = O($对数 $n)$ 且 $\mu = k^2/n$，我们有 $t \geqslant \mu - W = \Theta(\mu)$ 和 $s \leqslant D_n\sqrt{\mu} = o(\mu)$。因此，当 μ 足够大时，$(t+s+2)^2 \geqslant c\mu^2$，故存在某个普适常数 $C > 0$

$$|f'(t)| \leqslant \frac{C(s+1)}{\mu^2} \leqslant \frac{CD_n\sqrt{\mu}}{\mu^2} = \frac{CD_n}{\mu^{3/2}}.$$

由中值定理及 $|t-t'| \leqslant 2W$,

$$|r_t(s) - r_{t'}(s)| \leqslant \sup_{\xi\in\mathcal{I}} |f'(\xi)||t-t'| \leqslant \frac{CD_n}{\mu^{3/2}} \cdot (2W) = O\left(\frac{D_n\log\mu}{\mu}\right).$$

□

为了证明非典型取值 $t \notin \mathcal{I}$ 的贡献可以忽略，我们还需要分母 $\mathbf{P}(S = s_n)$ 的一个下界。下一个引理给出了相关尺度 $s_n = D_n\sqrt{\mu}$ 下的此类界。

引理 10.10. 假设 $k = n^{\beta+o(1)}, \beta \in (1/2, 1)$ 和 $s_n = \Theta\left(D_n\sqrt{\mu}\right)$ 满足 $D_n \leqslant C$对数 $n, C > 0$和$\mu = \omega($对数$^6\, n)$。那么对于所有充分大的 n,

$$\mathbb{P}(S = s_n) \geqslant \frac{c_1}{\mu}\exp\Big(-\frac{s_n^2(1+o(1))}{2\mu}\Big) \geqslant \frac{c_1}{\mu}\exp\Big(-\Theta(D_n^2)\Big),$$

其中 $c_1 > 0$ 为一个普适常数。

证明。 令 J:= $\{t:\ |t-\mu| \leqslant 4\sqrt{\mu}\}$。由于 $\mathbf{Var}(T) = \mu(1-q) \leqslant \mu$，由切比雪夫不等式可得 $\mathbf{P}(T \in J) \geqslant$15/16。此外 $\mathbb{E}[(-1)^T] = (1-2q)^n = (1-2\mu/n)^n \leqslant e^{-2\mu},$，因此对于任意一种奇偶性,

$$\mathbb{P}(T \equiv s_n \pmod 2) = \frac{1}{2}\Big(1 + (-1)^{s_n}\mathbb{E}[(-1)^T]\Big) \geqslant \frac{1}{2} - \frac{e^{-2\mu}}{2}.$$

因此，根据并集界,

$$\mathbb{P}(T \in J,\ T \equiv s_n \pmod 2) \geqslant 1 - \frac{1}{16} - \Big(\frac{1}{2} + \frac{e^{-2\mu}}{2}\Big) \geqslant \frac{1}{3}$$

对于所有足够大的 μ。令 $J' := \{t \in J:\ t \equiv s_n \ (\text{模}\ 2)\}$，并注意到在 J 中的整数里，至多有 $4\sqrt{\mu}+1$ 个具有 s_n奇偶性，因此存在 $t_{s_n} \in J'$ 使得

$$\mathbb{P}(T = t_{s_n}) \geqslant \frac{1/3}{|J'|} \geqslant \frac{1/3}{4\sqrt{\mu}+1} = \Omega(\mu^{-1/2}).$$

现在应用引理 10.7 并仅保留项 $t = t_{s_n}$:

$$\mathbb{P}(S = s_n) \geqslant \mathbb{P}(T = t_{s_n})\mathbb{P}(A_{t_{s_n}} = s_n).$$

由于 $t_{s_n} = \mu + O(\sqrt{\mu})$, 一对于较大的 μ, 我们有 $\overline{s}_n \leqslant D_n\sqrt{\mu} \leqslant 2D_n\sqrt{t_{s_n}}$。

现在回忆一下, $A_t = \sum_{i=1}^{t} \varepsilon_i$ 满足 $\varepsilon_i \overset{iid}{\sim} \mathrm{Rad}(\pm 1)$, 因此引理 10.1 可应用于 $n = t$, $T = D_n$ 且 $|s| \leqslant D_n\sqrt{t}$ 因为 $D_n = o(t^{1/6})$。事实上, 我们有 $|s| = o(t^{2/3})$, 因此特别地 $|s| \leqslant t/2$ 对所有充分大的 t均成立。因此, 对于所有较大的 t,

$$\mathbb{P}(A_t = s) \geqslant \frac{c_0}{\sqrt{t}} \exp\Big(-\frac{s^2}{2t}\Big) \tag{10.7}$$

其中 $c_0 > 0$是一个普适常数。等式 (10.7) 得出

$$\mathbb{P}(A_{t_{s_n}} = s_n) \geqslant \frac{c_0}{\sqrt{t_{s_n}}} \exp\Big(-\frac{s_n^2}{2t_{s_n}}\Big) \geqslant \frac{c_0'}{\sqrt{\mu}} \exp\Big(-\frac{s_n^2}{2\mu}(1 + o(1))\Big).$$

两边乘以 $\mathbb{P}(T = t_{s_n}) \geqslant c/\sqrt{\mu}$ 即可得到结果。 □

对于固定的整数 $s \in \mathbb{N}$, 定义整数 $t_\star = t_\star(s)$ 为满足下列条件的任意整数

$$t_\star \equiv s \mod 2, \qquad |t_\star - \mu| \leqslant 1. \tag{10.8}$$

现在, 我们可以结合混合变量 T的集中性、 $r_t(s)$ 在典型窗口上的稳定性, 以及 $\mathbb{P}(S = s_n)$的下界。这使我们能够近似计算 S的对数第二步比率。

引理 10.11。 假设 $k = n^{\beta+o(1)}, \beta \in (1/2, 1)$ 和 $D_n \leqslant C$对数 n, $C > 0$。设 $s = \Theta(D_n\sqrt{\mu})$。那么

$$\log \mathbb{P}(S = s_n + 2) - \log \mathbb{P}(S = s_n) = \log\Big(\frac{t_\star - s_n}{t_\star + s_n + 2}\Big) + O\Big(\frac{D_n \log n}{\mu}\Big).$$

证明。 由引理 10.8,

$$\frac{\mathbb{P}(S = s_n + 2)}{\mathbb{P}(S = s_n)} = \sum_{t=0}^{n} u_t(s_n) r_t(s_n), \qquad u_t(s_n) := \frac{w_t(s_n)}{\sum_u w_u(s_n)}.$$

将求和拆分为 $t \in \mathcal{I}$和 $t \notin \mathcal{I}$, 其中 $\mathcal{I} = \{|t - \mu| \leqslant W\}$, 且对于某个固定的 $C' > 0$, 有 $W = C'\sqrt{\mu}$对数 μ。步骤1 (在 $\mathcal{I}$ 之外): 由于 $0 \leqslant \mathbb{P}(A_t = s_n) \leqslant 1$,

$$\sum_{t \notin \mathcal{I}} w_t(s_n) \leqslant \mathbb{P}(T \notin \mathcal{I}) \quad \text{and therefore} \quad \sum_{t \notin \mathcal{I}} u_t(s_n) \leqslant \frac{\mathbb{P}(T \notin \mathcal{I})}{\mathbb{P}(S = s_n)}.$$

由标准的 Chernoff 界可得, 对于 $u \in (0, \mu)$, $\mathbb{P}(|T - \mu| > u) \leqslant 2\exp(-u^2/(3\mu))$。代入 $u = C'\sqrt{\mu}$对数 $\overline{\mu}$可得 $u^2/(3\mu) = \frac{C'^2}{3}$对数$^2\mu$。因此, $\mathbb{P}(T \notin \mathcal{I}) \leqslant 2e^{-c\log^2\mu}$, 其中 $c = C'^2/3$, 且由引理 10.10, $\mathbb{P}(S = s_n) \geqslant (c_1/\mu)$ $\exp(-s_n^2/(2\mu))$。因此

$$\sum_{t \notin \mathcal{I}} u_t(s_n) \leqslant \mu \cdot 2e^{-c\log^2\mu} \cdot \exp\Big(\frac{s_n^2}{2\mu}\Big) \cdot \frac{1}{c_1} = \exp\Big(-c\log^2\mu + \frac{s_n^2}{2\mu} + O(\log n)\Big).$$

由于 $s_n^2/\mu = \Theta(D_n^2)$, $\mu = n^{\Theta(1)}$ 和 $D_n \leqslant C$对数n, 选择足够大的 C':

$$\sum_{t \notin \mathcal{I}} u_t(s_n) = O(e^{-\log^2 n}). \tag{10.9}$$

步骤 2 (内部 $\mathcal{I}$): 根据引理 10.9,

$$\sup_{t \in \mathcal{I}} |r_t(s_n) - r_{t_\star}(s_n)| = O\Big(\frac{CD_n \log\mu}{\mu}\Big).$$

步骤 3（得出结论）：对于每个整数 $s \geqslant 0$ 以及每个满足 $\mathbb{P}(A_t = s) > 0$ 的 t（等价地，$s \equiv t$（模 2）且 $|s| \leqslant t$），

$$r_t(s) = \frac{\mathbb{P}(A_t = s+2)}{\mathbb{P}(A_t = s)} = \frac{\binom{t}{(t+s+2)/2}}{\binom{t}{(t+s)/2}} = \frac{t-s}{t+s+2}.$$

因此，对于 $s \geqslant 0$，有 $0 \leqslant r_t(s_n) \leqslant 1$，从而

$$\begin{aligned}\left|\sum_t u_t(s_n) r_t(s_n) - r_{t_\star}(s_n)\right| &\leqslant \sup_{t \in \mathcal{I}} |r_t(s_n) - r_{t_\star}(s_n)| + 2\sum_{t \notin \mathcal{I}} u_t(s_n) \\ &\leqslant \frac{C D_n \log n}{\mu} + o\left(\frac{D_n \log n}{\mu}\right),\end{aligned}$$

其中，对于最后一个不等式，我们使用了步骤 2 和（10.9）。因此，

$$\frac{\mathbb{P}(S = s_n + 2)}{\mathbb{P}(S = s_n)} = r_{t_\star}(s_n) + O\left(\frac{D_n \log n}{\mu}\right) = \frac{t_\star - s_n}{t_\star + s_n + 2} + O\left(\frac{D_n \log n}{\mu}\right).$$

记 $R := \frac{t_\star - s_n}{t_\star + s_n + 2}$，并令 $\delta = O\left(\frac{D_n \text{对数} n}{\mu}\right)$。由于 $t_\star = \mu + O(1)$ 且 $s_n = o(\mu)$，我们有 $R = \Theta(1)$，因此

$$\log \frac{\mathbb{P}(S = s_n + 2)}{\mathbb{P}(S = s_n)} = \log(R + \delta) = \log R + \log\left(1 + \frac{\delta}{R}\right) = \log R + O\left(\frac{\delta}{R}\right) = \log R + O\left(\frac{D_n \log n}{\mu}\right),$$

这就证明了我们所需的结果。取对数（利用主比值远离 0 和 8 这一事实，因为 $s_n = o(\mu)$），即可得到所声称的形式。□

现在已将分子确定到一个可控误差范围内，接下来只需除以增量$(s_n + 2)^r - s_n^r$。下面的命题完成最后这一步展开，并给出所声称的 $\Lambda_r(s_n)$ 阶。

命题 10.12。固定一个整数 $r \geqslant 2$。假设 $k = n^{\beta + o(1)}$、$\beta \in (1/2, 1)$ 且 $D_n \leqslant C$对数 n。令 $s_n = \Theta(D_n \sqrt{\mu})$。那么，

$$\Lambda_r(s_n) = -\Theta\left(\frac{1}{D_n^{r-2} \mu^{r/2}}\right).$$

证明。根据引理 10.11，

$$\log \mathbb{P}(S = s+2) - \log \mathbb{P}(S = s) = \log\left(\frac{t_\star - s_n}{t_\star + s_n + 2}\right) + O\left(\frac{D_n \log n}{\mu}\right).$$

应用引理 10.6，其中 $t = t_\star$定义于（10.8）中，且 $s = s_n$：

$$\log\left(\frac{t_\star - s_n}{t_\star + s_n + 2}\right) = -\frac{2s_n + 2}{t_\star} + O\left(\frac{(s_n+1)^2}{t_\star^2}\right).$$

结合 $t_\star = \mu + O(1)$，

$$\log \mathbb{P}(S = s+2) - \log \mathbb{P}(S = s) = -\frac{2s_n + 2}{\mu} + O\left(\frac{(s_n+1)^2}{\mu^2}\right) + O\left(\frac{D_n \log n}{\mu}\right). \tag{10.10}$$

现在除以 $(s_n + 2)^r - s_n^r$。由于 $s_n \to \infty$，当 $D_n \sqrt{\mu} \to \infty$ 时，利用二项式定理：

$$(s+2)^r - s^r = \sum_{j=1}^{r} \binom{r}{j} 2^j s^{r-j} = 2rs^{r-1} + \sum_{j=2}^{r} \binom{r}{j} 2^j s^{r-j} = 2rs^{r-1} + O_r(s^{r-2}).$$

因此，

$$(s_n + 2)^r - s_n^r = 2rs_n^{r-1}\left(1 + O_r\left(\tfrac{1}{s_n}\right)\right)$$

这意味着

$$\frac{1}{(s_n+2)^r - s_n^r} = \frac{1}{2rs_n^{r-1}}\left(1 + O_r\left(\tfrac{1}{s_n}\right)\right). \tag{10.11}$$

将（10.10）、（10.11）相乘可得

$$\Lambda_r(s_n) = -\frac{1}{r\mu} s_n^{-(r-2)}\left(1 + O_r\left(\tfrac{1}{s_n}\right)\right) + O\left(\frac{D_n \log n}{\mu s_n^{r-1}}\right) + O\left(\frac{(s_n+1)^2}{\mu^2 s_n^{r-1}}\right). \tag{10.12}$$

特别　　地，由于 $s_n = \Theta(D_n\sqrt{\mu})$ 这就得到了所需的不等式，证明完成　　e.　　□

10.1.4 引理 10.4的证明

设 $X_i := v_i v_i' \in \{-1,0,+1\}, i \in [n]$ 且 $T := \sum_{i=1}^n \mathbf{1}\{|X_i| = 1\}$。那么 $T \sim$ 二项分布(n,q) 满足 $q = (k/n)^2$ 和 $\mathbb{E}[T] = \mu = nq$。给定 $\{T = t\}$的条件下， t 非零的 X_i 独立同分布且在 $\{\pm 1\}$ 上均匀分布，因此 $S_n \mid (T = t) \overset{d}{=} \sum_{j=1}^t \xi_j$ 满足 $\xi_j \overset{iid}{\sim} \mathrm{Rad}(\pm 1)$。特别地， $S_n \leqslant T$ 始终成立，而在 $\{T = m\}$ 上有 $\mathbf{P}(S_n \geqslant m \mid T = m) = \mathbf{P}(|\sum_{j=1}^m \xi_j| = m) = 2^{1-m}$。因此，对于每个整数 $m \geqslant 1$,

$$2^{1-m}\,\mathbb{P}(T = m) \;\leqslant\; \mathbb{P}(S_n \geqslant m) \;\leqslant\; \mathbb{P}(T \geqslant m). \tag{10.13}$$

注意到

$$\mathbb{P}(T \geqslant m) \leqslant \binom{n}{m} q^m \leqslant \frac{n^m}{m!}\, q^m = \frac{(nq)^m}{m!} = \frac{\mu^m}{m!}. \tag{10.14}$$

接下来，对于 $m = n^{o(1)}$，且由于 $\mu = o(1)$,

$$\mathbb{P}(T = m) = \binom{n}{m} q^m (1-q)^{n-m} = \frac{\mu^m}{m!}\,(1 + o(1)),$$

由于 $\binom{n}{m} = n^m/m! \cdot (1 + o(1))$ 且 $(1-q)^{n-m} = \exp(-(n-m)q + o(1)) = 1 + o(1)$。将其与（10.13）和（10.14）结合，可得对于 $m = n^{o(1)}$,

$$\frac{\mu^m}{m!}\, 2^{-m} \;\leqslant\; \mathbb{P}(S_n \geqslant m) \;\leqslant\; \frac{\mu^m}{m!}.$$

取对数并利用 $\log m! = m\log m - m + O(\log m)$，得到

$$-\log \mathbb{P}(S_n \geqslant m) = m \log \frac{1}{\mu} + O(m \log m).$$

由于 $\log(1/\mu) = (1-2a)\log n$，项 $m \log m$ 相对于 $m\log(1/\mu) = \Theta(m \log n)$ 是低阶的，因此存在常数 c, $C > 0$，使得对所有整数 $c_0 \log n \leqslant D_n \leqslant n^{o(1)}$,

$$-\log \mathbb{P}(S_n \geqslant m) \geqslant D_n \;\text{ for all }\; m \geqslant C\,\frac{D_n}{\log n}, \qquad -\log \mathbb{P}(S_n \geqslant m) \leqslant D_n \;\text{ for all }\; m \leqslant c\,\frac{D_n}{\log n}.$$

根据 $q(D_n)$ 的定义以及 $m \mapsto \mathbf{P}(S_n \geqslant m)$ 的单调性，这意味着

$$c\,\frac{D_n}{\log n} \;\leqslant\; q(D_n) \;\leqslant\; C\,\frac{D_n}{\log n},$$

如所述。

10.1.5 引理10.5 的证明

令 $S = \langle v,\ v' \rangle$，于是 $\langle X,\ X' \rangle = S^r$。

我们首先断言，对所有满足 $m = n^{o(1)}$ 的整数 $m \geqslant 1$，该结论均成立,

$$\mathbb{P}(S = m) = \frac{\mu^m}{2^m m!}\,(1 + o(1)), \tag{10.15}$$

其中，与之前一样， $\mu = k^2/n \to 0$。事实上，由引理10.7，

$$\mathbb{P}(S = m) = \sum_{j \geqslant 0} \mathbb{P}(T = m + 2j)\, \mathbb{P}(A_{m+2j} = m),$$

其中 $T \sim$ 二项分布$(n,\ q)$ 且 $A_t = \sum_{i=1}^{t} \xi_i$ 满足 $\xi_i \overset{iid}{\sim} \mathsf{Rad}(\pm 1)$。对于 $j = 0$ 项，

$$\mathbb{P}(T = m)\, \mathbb{P}(A_m = m) = \mathbb{P}(T = m)\, 2^{-m}.$$

由于 $m = n^{o(1)}$，我们有一致地

$$\mathbb{P}(T = m) = \binom{n}{m} q^m (1-q)^{n-m} = \frac{\mu^m}{m!}\,(1 + o(1)).$$

因此

$$\mathbb{P}(T = m)\, \mathbb{P}(A_m = m) = \frac{\mu^m}{2^m m!}\,(1 + o(1)).$$

剩下要证明 $j \geqslant 1$ 的贡献可以忽略。利用对所有 $t \geqslant 0$都有 $\mathbb{P}(T = t) \leqslant \mu^t/t!$，可得

$$\begin{aligned}
\sum_{j \geqslant 1} \mathbb{P}(T = m + 2j)\, \mathbb{P}(A_{m+2j} = m) &\leqslant \sum_{j \geqslant 1} \frac{\mu^{m+2j}}{(m+2j)!} \binom{m+2j}{j} 2^{-(m+2j)} \\
&= \frac{\mu^m}{2^m} \sum_{j \geqslant 1} \frac{\mu^{2j}}{4^j\, j!\, (m+j)!} \\
&\leqslant \frac{\mu^m}{2^m m!} \sum_{j \geqslant 1} \frac{\mu^{2j}}{4^j\, j!} = O\left(\frac{\mu^{m+2}}{2^m m!}\right).
\end{aligned}$$

由于 $\mu \to 0$，这 $o(\mu^m/(2^m m!\ ))$ 关于 $m = n^{o(1)}$一致成立。这证明了（10.15）。

将（10.15）应用于 $m = s_n$ 和 $m = s_n + 2$，得到

$$\frac{\mathbb{P}(S = s_n + 2)}{\mathbb{P}(S = s_n)} = \frac{\mu}{2(s_n + 2)}\,(1 + o(1)),$$

因此

$$\log \mathbb{P}(S = s_n) - \log \mathbb{P}(S = s_n + 1) = \log \frac{1}{\mu} + \log(2(s_n + 1)) + o(1). \tag{10.16}$$

现在 $\mu = n^{-(1-2\beta)+o(1)}$，所以对数 $\mu = (2\beta - 1 + o(1))$对数 n. 此外，由于 $s_n = n^{o(1)}$，对数$(s_n + 2) = o($对数 $n)$. 因此 (10.16) 得到

$$\log \mathbb{P}(S = s_n) - \log \mathbb{P}(S = s_n + 2) = \Theta(\log n). \tag{10.17}$$

接下来，由于 $s_n \geqslant 1$ 和 $r \geqslant 2$，我们有 $(s_n + 1)^r - s_n^r = \Theta(s_n^{r-1})$.

因此，

$$-\Delta_{a_n} \log \mathbb{P}(\langle X, X' \rangle = s_n^r) = \frac{\log \mathbb{P}(S = s_n) - \log \mathbb{P}(S = s_n + 2)}{(s_n + 2)^r - s_n^r} = \Theta\left(\left(\frac{\log n}{(s_n)^{r-1}}\right)\right),$$

这就证明了引理的第一部分。

现在令 $s_n :=$ $q_n(D_n)$. 由引理 10.4， $s_n = \Theta\,(D_n/$对数 $n)$. 由于 $D_n = n^{o(1)}$ 且 $D_n \geqslant c_0$ 对数 n，因此 $1 \leqslant s_n = n^{o(1)}$ 对于所有充分大的 n。此外，

$$(s_n + 1)^r - s_n^r = \Theta\left(\left(\frac{D_n}{\log n}\right)^{r-1}\right). \tag{10.18}$$

因此，将上述结果结合起来，对于这一选择的 s_n，

$$-\Delta_{a_n} \log \mathbb{P}(\langle X, X' \rangle = s_n^r) = \Theta\left(\frac{\log n}{(D_n/\log n)^{r-1}}\right) = \Theta\left(\frac{(\log n)^r}{D_n^{r-1}}\right).$$

这就证明了该引理。

10.2 张量主成分分析的证明

设 $v,\ v' \sim \mathcal{N}(0,\ I_n)$ 独立抽取，并设 $S = \langle v,\ v' \rangle$。那么， $X = v^{\otimes r},\ X' = (v')^{\otimes r}$ 是高斯张量主成分分析中的两次独立同分布抽样，且 $S^r = \langle X,\ X' \rangle$。另外，记 f_S 为 S的密度。下面给出证明定理 3.1 所需用到的两个引理。

如前所述，对于每个模型，这些引理要完成以下工作：求出分位数的主导阶 $q(D_n)$，然后求出式（2.6）中导数在阶数$q(D_n)$处的阶。这些步骤使我们能够将定理2.7应用于张量主成分分析。

引理 10.13。 固定一个整数 $r \geqslant 2$。那么存在绝对常数 $0 < c < C < \infty$ 和一个绝对常数 $n_0 \geqslant 1$，使得对于所有 $n \geqslant n_0$ 和所有 $\Omega(1) = D \leqslant n$,

$$\left(c(\sqrt{nD} + D)\right)^r \leqslant q(D) \leqslant \left(C(\sqrt{nD} + D)\right)^r.$$

该引理的证明推迟到第 10.2.2 节。

引理 10.14。 对于具有高斯先验的张量 *PCA* 模型，任意 $1 \leqslant D_n = o(\sqrt{n})$ 对应的导数阶为：

$$-\frac{d}{ds}\log f_S(s)\Big|_{s=(D_n n)^{r/2}} = \Theta(n^{-r/2} D_n^{(3-r)/2}) = \Theta(\lambda_{\mathrm{ALG}} \cdot D_n^{(3-r)/2}).$$

该引理的证明推迟到第 10.2.2 节。

10.2.1 定理 3.1 的证明

我们通过直接应用定理 2.7 来证明这一点。为此，我们首先证明先验分布满足主要假设2.4。具体来说，我们验证这四个条件。

条件 1： 由于标准高斯分布的所有阶累积量均为非负，且张量主成分分析具有（8.1）的形式，引理 8.5 表明，对于所有多重指标 $\alpha \in \mathbf{N}^n$，均有 $\kappa_\alpha(X_i) \geqslant 0$.条件 2： 这直接源于以下事实：每个 X_i 都是 r 高斯随机变量的乘积。条件 3： 该条件显然成立，因为对于 $D'_n = \Theta(D_n \text{对数}^2\, n)$，并利用引理 10.13

$$q(D'_n) \geqslant (cD'_n)^r = \omega(\max\{\log^{-D_n} n, n^{-C}\})$$

对于任意 $C > 0$。

条件 4： 利用引理 10.13，我们知道对于 $D_n = o(\sqrt{n})$，在 $q(D_n)$ 处的导数 $(nD_n)^{r/2}$.为 。因此，假设 2.4 中的第 4 项等价于证明存在一个常数 $C > 0$ 使得:

$$\frac{D_n^{(3-r)/2} n^{-r/2}}{A_n} \leqslant \frac{1}{q(D'_n)\log n}.$$

根据引理 10.13 我们知道，对于 $D'_n \leqslant \sqrt{n}$，有

$$q(D'_n) \leqslant C_1 (D'_n n)^{r/2}.$$

代入 将这个不等式代入上面的不等式，可得假设 4 对所有 $D'_n = o(\sqrt{n})$ i 都成立 f:

$$\frac{D_n^{(3-r)/2} n^{-r/2}}{A_n} \leqslant \frac{1}{C'_1 n^{r/2} (D'_n)^{r/2} \log n} \quad \text{or equivalently} \quad C'_1 D_n^{3/2} \log^{r+1} n \leqslant A_n$$

这里我们使用了 $D'_n = D_n$对数$^2\, n$，得到最后一个不等式。因此，对于 $A_n = C'_1 (D_n)^{3/2}$ 对数$^{r+1}\, n$，该不等式成立。

现在将上述结果与（2.6）结合起来，我们得到对所有 $\lambda > 0$

$$\mathrm{Corr}_{P_0}^{\leqslant D_n}\left(\frac{1}{C'_1 D_n^{3/2}\log^{r+1} n}\cdot\left(\lambda + \frac{\mathrm{d}}{\mathrm{d}q}\mathcal{F}_{\mathrm{ann},\lambda}\Big|_{q=q(D_n)}\right)\right)^2 \leqslant q(D'_n) \leqslant C_1 (D'_n n)^{r/2}.$$

现在注意到

$$\mathrm{MMSE}_X^{\mathrm{trivial}} = \mathbb{E}\|X\|^2 - (\mathbb{E}\|X\|)^2 = n^r. \tag{10.19}$$

此外，由引理 10.14 和 (2.6) 对所有 $\lambda > 0$，$\lambda + \mathrm{d}\,\mathrm{d}q\mathcal{F}_{\mathrm{ann},\lambda}\Big|_{q=q(D_n)} = \Theta(D_n^{(3-r)/2}\lambda_{\mathrm{ALG}})$。取 $\lambda = D_n^{(3-r)/2}n^{-r/2}/$对数 n，现在可得 $\mathrm{d}\,\mathrm{d}q\mathcal{F}_{\mathrm{ann},\lambda}\Big|_{q=q(D_n)} \geqslant 0$。因此，根据定理 2.7，相关性$(\lambda/A_n)^2 \leqslant 2q(D_n')$，其中 $A_n = C_1'(D_n)^{3/2}$对数$^{r+1}\, n$。等价地，

$$\mathrm{MMSE}_X^{\leqslant D_n}\left(\frac{\lambda}{C_1' D_n^{3/2}\log^{r+1} n}\right) \geqslant n^r - C_1(D_n' n)^{r/2}.$$

最后，代入 $\lambda_{\mathrm{ALG}} = n^{-r/2}$

$$\mathrm{MMSE}_X^{\leqslant D_n}\left(\frac{\lambda_{\mathrm{ALG}}}{C_1' D_n^{r/2}\log^{r+2} n}\right) \geqslant n^r - C_1(D_n' n)^{r/2}.$$

将其与 (10.19) 结合，我们证明了对于任意 $D_n' = o(\sqrt{n})$ 和 $D_n = {}^{D_n'}$对数$^2\, n = o(\sqrt{n}/$对数$^2\, n) -$

$$\mathrm{MMSE}_X^{\leqslant D_n}\left(\frac{\lambda_{\mathrm{ALG}}}{C_1' D_n^{r/2}\log^{r+2} n}\right) \geqslant \mathrm{MMSE}_X^{\mathrm{trivial}}(1+o(1)),$$

从而完成证明。

10.2.2 引理的证明 10.13

上界：$X_i := v_i v_i'$, $i \in [n]$ 是独立同分布的零均值次指数变量，因此存在某个绝对常数 $c_0 > 0$,

$$\mathbb{P}(|S| \geqslant t) \leqslant 2\exp\left(-c_0 \min\left\{\frac{t^2}{n}, t\right\}\right) \qquad t \geqslant 0,$$

这意味着当绝对常数 C 足够大时，$\mathbb{P}(|S| \geqslant C(\sqrt{nD}+D)) \leqslant e^{-D}$。因此 $q(D) \leqslant (C(\sqrt{nD}+D))^r$。下界：对 v 条件化。由于 $v' \sim \mathcal{N}(0, I_n)$ 独立于 v,

$$S = \langle v, v'\rangle \mid v \sim \mathcal{N}\big(0, \|v\|_2^2\big).$$

令 $U := \|v\|_2^2 \sim \chi_n^2$，并令 $Z \sim \mathcal{N}(0,1)$ 独立于 U。于是 $S \overset{d}{=} \sqrt{U}Z$。令 $E := \{U \geqslant n/2\}$。由于$\mathbb{E}U = n$ 且 $\mathsf{Var}(U) = 2n$，由切比雪夫不等式可得

$$\mathbb{P}(E) \geqslant 1 - \frac{\mathrm{Var}(U)}{n^2} = 1 - \frac{8}{n} \geqslant \frac{1}{2},$$

对所有 $n \geqslant 8$ 均成立。在 E 上，我们有 $\sqrt{U} \geqslant \sqrt{n}/2$，，因此对任意 $t > 0$,

$$\mathbb{P}(|S| \geqslant t) \geqslant \mathbb{P}(E)\mathbb{P}\left(|Z| \geqslant \frac{t}{\sqrt{n/2}}\right) \geqslant \frac{1}{2}\mathbb{P}\left(|Z| \geqslant \frac{t}{\sqrt{n/2}}\right).$$

选择较小的 $t = c\sqrt{nD}$，使得 $c > 0$，因此 $t/\sqrt{n}/2 = c\sqrt{D}/2$。利用标准高斯尾部下界，存在一个常数 $c_1 > 0$, 使得对于 $x \geqslant c_1$,

$$\mathbb{P}(|Z| \geqslant x) \geqslant \frac{c_1^2}{3(c_1^2+1)x} e^{-x^2/2},$$

由于 $D \geqslant 1$，我们有 $x = \Theta(\sqrt{D}) \geqslant c_1$。因此，对于一个只依赖于 c_1 的绝对常数 c_2，

$$\mathbb{P}(|S| \geqslant c\sqrt{nD}) \geqslant \frac{c_2}{\sqrt{D}}\exp\left(-\frac{c^2}{4}D\right) \geqslant e^{-D},$$

对所有 $\Omega(1) = D \leqslant n$，取足够大的 $0 < c < 2$ 和 $D \geqslant c_2^2$。因此，在此情形下，$|S|$ 的 e^{-D} 分位数至少为 $c\sqrt{nD}$，从而 $q(D) \geqslant (c(\sqrt{nD}+D))^r$。

10.2.3 引理的证明 10.14

假设 v, $v' \sim \mathcal{N}(0, I_n)$ 相互独立。于是，正如引理 9.5所证明的， $T = \langle v, v' \rangle$ 的密度可以用第二类修正贝塞尔函数 K_ν显式表示:

$$f_T(t) = \frac{1}{\sqrt{\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(\frac{|t|}{2}\right)^{\frac{n}{2}-\frac{1}{2}} K_{\frac{n}{2}-\frac{1}{2}}(|t|), \qquad t \in \mathbb{R}.$$

因此,

$$\log f_T(t) = C + \left(\frac{n}{2} - \frac{1}{2}\right) \log t + \log K_{\frac{n}{2}-\frac{1}{2}}(t),$$

对于常数 C ，其与 t 无关，引理 9.1 给出

$$\frac{\mathrm{d}}{\mathrm{d}t} \log \mathbb{P}(\langle v, v' \rangle = t) = \frac{\mathrm{d}}{\mathrm{d}t} \log f_T(t) = \frac{n-1}{t} - \frac{K_{(n+1)/2}(t)}{K_{(n-1)/2}(t)}.$$

令 $\nu = \frac{n-1}{2}$。回顾引理9.1 中陈述的第二类修正贝塞尔函数的恒等式，另见 [AS65，第 9],

$$K_{\nu+1}(t) = \frac{2\nu}{t} K_\nu(t) + K_{\nu-1}(t), \qquad t > 0.$$

因此,

$$\frac{\mathrm{d}}{\mathrm{d}t} \log \mathbb{P}(\langle v, v' \rangle = t) = \frac{2\nu}{t} - \frac{K_{\nu+1}(t)}{K_\nu(t)} = -\frac{K_{\nu-1}(t)}{K_\nu(t)},$$

对所有 $t > 0$。根据推论 9.4，该推论由 [YC17 推导得出，(1.10)], 对于 $t = \sqrt{nD_n}$ 以及 $\overline{D_n} = o(\sqrt{n})$,

$$\frac{K_{\nu-1}(t)}{K_\nu(t)} = \frac{t}{2\nu}(1 + o(1)) = \sqrt{\frac{D_n}{n}}(1 + o(1)).$$

然后，根据链式法则，如果 $t = s^{1/r}$,

$$\frac{\mathrm{d}}{\mathrm{d}s} \log \mathbb{P}(\langle X, X' \rangle = s) = \frac{\mathrm{d}}{\mathrm{d}s} \log \mathbb{P}(\langle v, v' \rangle^r = s) = \frac{\mathrm{d}}{\mathrm{d}t} \log \mathbb{P}(\langle v, v' \rangle = t) \cdot \frac{t^{1-r}}{r},$$

因此代入 $t = \sqrt{nD_n}$（即 $s = (nD_n)^{r/2}$）可得

$$-\frac{\mathrm{d}}{\mathrm{d}s} \log \mathbb{P}(\langle X, X' \rangle = s)\Big|_{s=(nD_n)^{r/2}} = \Theta\left(\sqrt{\frac{D_n}{n}}\right) \cdot \frac{(nD_n)^{(1-r)/2}}{r} = \Theta\left(n^{-r/2} D_n^{1-r/2}\right).$$

10.3 稀疏聚类的证明

令 $\xi = (\xi_1, \ldots, \xi_n)^\top$ 具有独立同分布的 Rad 分布$(1/2)$ 的元素，并注意到，该模型中的信号矩阵可以写成 $X := \xi\mu^\top \in \mathbb{R}^{n\times p}$，即 $X_{ij} = \xi_i \mu_j$。正如我们所讨论的，当将其写成形式为（1.1）的高斯加性模型时，信噪比 $\lambda = \Delta/s$将为 $(\xi_i)_{i=1}^n$ （经过全局符号翻转后）从 Y中恢复标签。

10.3.1 符号说明

我们先介绍一些符号说明。令 $n, p \geqslant 1$ ，其中 $p = n^{c+o(1)}$，对某个常数 $c > 0$，且 $s = s_{n,p}$满足 $\widetilde{\omega}(\sqrt{p}) = s = o(p)$。再令 $\xi_i, \xi_i' \overset{iid}{\sim} \mathsf{Rad}(\pm 1)$ ，其中 $i \in [n]$。定义

$$A := \sum_{i=1}^n \xi_i \xi_i'.$$

则 $\mathbb{E}[A]=0$，且 $\mathrm{Var}(A)=n$。令 $b_j, b'_j \overset{iid}{\sim} \mathrm{Ber}(s/p)$ 和 $g_j, g'_j \overset{iid}{\sim} N(0,1)$，它们彼此独立。于是，我们可以定义 μ, μ' 为从先验中独立抽取的两个样本

$$\mu_j := b_j g_j,\ \mu'_j := b'_j g'_j$$

并记 $Z_j := \mu_j \mu'_j = (b_j b'_j)(g_j g'_j)$, $B := \sum_{j=1}^p Z_j$。再令 $\delta_j := b_j b'_j \sim \mathrm{Ber}((s/p)^2)$ 和 $Y_j := g_j g'_j$。那么，$Z_j = \delta_j Y_j$ 和 $U := \sum_{j=1}^p \delta_j \sim$ 二项分布$(p,\ (s/p)^2)$. 使用这一符号说明，若令 $S = \langle X,\ X' \rangle$ 则

$$S = A \cdot B.$$

特别地，$\mathrm{Var}(S) = n\sigma_B = \sigma_S^2$。由于 B 在 $\mathbb{R}\backslash\{0\}$上连续，$S$ 在 $\mathbb{R}\backslash\{0\}$上具有绝对连续的密度。我们用 f_S 表示该密度在 $\mathbb{R}\backslash\{0\}$上的取值。最后，记 $\phi(x) := (2\pi)^{-1/2} e^{-x^2/2}, x \in \mathbb{R}$。

10.3.2 辅助引理

在证明定理3.3 之前，我们先陈述两个将要用到的引理。

再次，这些引理将发挥两方面的作用：找出式（2.6）中在阶数 $q(D_n)$ 处求值的导数的阶数，并计算分位数 $q(D_n)$ 的主阶。

引理 10.15。存在绝对常数 $0 < c < C < \infty$ 以及绝对常数 $n_0, \sigma_0 \geqslant 1$，使得只要$n \geqslant n_0$ 且 $\sigma_B \geqslant \sigma_0$，对所有 $1 \leqslant D \leqslant c\min\{n, \sigma_B\}$，均有

$$c\sigma_S D \leqslant q(D) \leqslant C\sigma_S D,$$

其中 σ_S^2 =方差(S)。

此引理的证明推迟至第 10.3.4 节。

引理 10.16。假设 $p \geqslant s = \tilde{\omega}(\sqrt{p})$。$S$ 的密度的对数导数对所有 D_n =$o(\max\{n^{1/3}, \sigma_B^{1/6}\})$ 满足：

$$\frac{d}{ds}\log f_S(s)\Big|_{s=D_n\sigma_S} = -\frac{1}{s}\sqrt{\frac{p}{n}} \cdot \big(1 + o(1)\big).$$

此引理的证明推迟至第 10.3.5 节。

10.3.3 定理3.3的证明

我们再次通过直接应用定理 2.7 来证明这一点。为此，我们首先证明所选先验分布满足主要假设 2.4。具体而言，我们验证这四个条件。

条件 1：由于该模型的形式为（8.1）（此处 $Z_j = \xi_j$，$t = n$ 以及 $a(i) =$ 所属的行）且我们赋予的先验分布 P_0 满足假设8.1（它是标准高斯分布与稀疏伯努利分布的乘积，因此均值为零），根据引理 8.2 可知，对于所有满足 $\alpha \in \mathbb{N}^n$ 且 $|\alpha| \leqslant D_n$对于 $D_n = \lfloor$对数$_2(p/s) - 1\rfloor$ 的多重指标，都有：$\kappa_\alpha(X_i) \geqslant 0.$

条件 2：这直接源于 X_i 是次高斯分布。

条件 3：第 假设的 rd 条件显然成立，因为对于 $D'_n = D_n \log^2 n$ 和 usin g 引理 10.15

$$q(D'_n) \geqslant cD'_n s\sqrt{n/p} = \omega(\max\{\log^{-D_n} n, n^{-C}\}),$$

对于某个普适正常数 $C > 0$，因为 $p = n^{c+o(1)}$，对于某个 $c > 0$。

条件 4：根据引理 10.15 我们知道，$q(D_n)$ 处的导数阶为 $\frac{D_n}{s}\sqrt{\frac{p}{n}}$. 因此，假设 2.4 中的第 4 项等价于以下条件：

$$\frac{\frac{1}{s}\sqrt{\frac{p}{n}}}{A_n} \leqslant \frac{1}{q(\log^2 n)\log n}.$$

根据引理 10.15 我们知道，对于 $D'_n = \Theta(\text{对数}^3\, n) \leqslant \mu$，有

$$q(D'_n) \leqslant C_1 D'_n s\sqrt{\frac{n}{p}}.$$

将此不等式代入上面的不等式，可得假设 4 对于 $A_n = C'_1 \log^4 n$ 成立，其中 $C'_1 > 0$ 是一个绝对常数。

现在将以上结果与 (2.6) 结合，可得对所有 $\lambda > 0$:

$$\mathrm{Corr}_{P_0}^{\leqslant D_n}\left(\frac{1}{C'_1 \log^4 n}\cdot\left(\lambda + \frac{\mathrm{d}}{\mathrm{d}q}\mathcal{F}_{\mathrm{ann},\lambda}\Big|_{q=q(D_n)}\right)\right)^2 \leqslant q(D'_n) \leqslant C_1 \frac{D'_n}{s}\sqrt{\frac{p}{n}}.$$

其中对于 最后一个不等式中，我们再次使用了不等式 $q(D'_n) \leqslant C_1 D'_n s\sqrt{\frac{n}{p}}$。现在注意到 t

$$\mathrm{MMSE}_X^{\mathrm{trivial}} = \mathbb{E}\|X\|^2 - (\mathbb{E}\|X\|)^2 = ns.$$

此外，由引理 10.16 和 (2.6)，对于所有 $\lambda > 0$，$\lambda + \mathrm{d}\,\mathrm{d}q\mathcal{F}_{\mathrm{ann},\lambda}\Big|_{q=q(D_n)} = \Theta(\frac{1}{s}\sqrt{p/n})$ 现在取$\lambda = {}_s{}^1$ $\log n\sqrt{p/n}$可推出，$\mathrm{d}\,\mathrm{d}q\mathcal{F}_{\mathrm{ann},\lambda}\Big|_{q=q(D_n)} \geqslant 0$。因此，根据定理 2.7，$\mathrm{Corr}(\lambda/A_n)^2 \leqslant 2q(D'_n)$,其中 $A_n = C'_1 \log^4 n$。等价地，

$$\mathrm{MMSE}_X^{\leqslant D_n}\left(\frac{\lambda}{C'_1 \log^4 n}\right) \geqslant ns - C_1 s \log^4 n\sqrt{\frac{n}{p}}.$$

利用 $\lambda_{\mathrm{ALG}} = \sqrt{p/(s^2 n)}$ （或等价地， $\Delta_{\mathrm{ALG}} = \sqrt{p/n}$）可得

$$\mathrm{MMSE}_X^{\leqslant D_n}\left(\frac{\lambda_{\mathrm{ALG}}}{C'_1 \log^5 n}\right) \geqslant ns - C_1 s \log^4 n\sqrt{\frac{n}{p}},$$

结合前两个不等式以及 $n = \omega(\log^8 n)$ 和 $p \geqslant 1$ 这一事实，我们证明了

$$\mathrm{MMSE}_X^{\leqslant D_n}\left(\frac{\lambda_{\mathrm{ALG}}}{C'_1 \log^5 n}\right) \geqslant (1 + o(1))\mathrm{MMSE}_X^{\mathrm{trivial}},$$

证毕。

10.3.4 引理10.15 的证明

我们在范围 $1 \leqslant D \leqslant c\min\{n, \sigma_B\}$ 内证明了关于 $q(D)$ 的匹配上界和下界。步骤1：根据 Hoeffding不等式，对所有 $t \geqslant 0$,

$$\mathbb{P}(|A| \geqslant t) \leqslant 2\exp\Big(-\frac{t^2}{2n}\Big).$$

反过来，根据引理10.1 存在一个普适常数 $c_A > 0$ ，使得对所有 n 足够大且对所有$1 \leqslant D \leqslant cn$,

$$\mathbb{P}\big(|A| \geqslant c_A\sqrt{nD}\big) \geqslant e^{-D},$$

类似地，存在一个普适常数 $C_A > 0$ 使得

$$\mathbb{P}\big(|A| \geqslant C_A\sqrt{nD}\big) \leqslant e^{-D}.$$

我们将指数中的任意固定常数吸收进去，因为可以调整 c_A，C_A。
步骤 2：现在我们以 (b, b', g') 为条件。$B = \sum_{j=1}^p (b_j b'_j g'_j)\, g_j$ 是独立$\mathcal{N}(0,1)$ 变量的线性组合，因此

$$B\big|(b, b', g') \sim \mathcal{N}(0, V), \qquad V := \sum_{j=1}^p (b_j b'_j g'_j)^2 = \sum_{j=1}^p b_j b'_j (g'_j)^2.$$

特别地，由于 $Z \sim \mathcal{N}(0,1)$ 与 V 相互独立，我们有精确表示 $B \overset{d}{=} \sqrt{V}\, Z$. 此外 $\mathbb{E}[V] = \sigma_B$ 由于 $\mathbb{E}[b_j b'_j] = (s/p)^2$ 和 $\mathbb{E}[(g'_j)^2] = 1$。

步骤 3：令 $Y_j := b_j b'_j (g'_j)^2 \geqslant 0$，于是 $V = \sum_{j=1}^{p} Y_j$，其中各项独立同分布。由于 $(g'_j)^2$ 的矩母函数为 $\mathbb{E}e^{\theta (g'_j)^2} = (1-2\theta)^{-1/2}$（$\theta < 1/2$,），对于 $\theta \in (0,\ 1/4)$

$$\mathbb{E}e^{\theta Y_j} = (1-(s/p)^2) + (s/p)^2 (1-2\theta)^{-1/2} \leqslant \exp\Big((s/p)^2 \big((1-2\theta)^{-1/2} - 1 \big) \Big),$$

它是有限的，且 对于固定的 $\theta < 1/2$，它关于 p 一致有界。对于任意 $\theta \in (0,\ 1/2),$，由切尔诺夫不等式可得 ，

$$\mathbb{P}(V \geqslant 2\sigma_B) \leqslant e^{-2\theta\sigma_B} \big(\mathbb{E}[e^{\theta Y_1}] \big)^p.$$

取 $\theta = \frac{1}{4}$，可得 $\mathbb{P}(V \geqslant 2\sigma_B) \leqslant e^{-c_+ \sigma_B}$，其中某个普适常数为 $c_+ > 0$。类似地，对于任意 $\lambda > 0$,

$$\mathbb{P}\Big(V \leqslant \frac{1}{2}\sigma_B \Big) = \mathbb{P}\Big(e^{-\lambda V} \geqslant e^{-\lambda\sigma_B/2} \Big) \leqslant e^{\lambda\sigma_B/2} \big(\mathbb{E}[e^{-\lambda Y_1}] \big)^p,$$

因此

$$\mathbb{P}\Big(V \leqslant \frac{1}{2}\sigma_B \Big) \leqslant \exp\Big(-\sigma_B \Big[1 - (1+2\lambda)^{-1/2} - \lambda/2 \Big] \Big).$$

取 $\lambda = \frac{1}{2}$，可得

$$\mathbb{P}\Big(V \leqslant \frac{1}{2}\sigma_B \Big) \leqslant e^{-c_- \sigma_B}$$

对于某个绝对常数 $c_- > 0$。

因此，对于某个通用常数 $c_V > 0$，$\mathbb{P}(V \geqslant 2\sigma_B) \leqslant e^{-c_V \sigma_B}$，$\mathbb{P}\Big(V \leqslant \frac{1}{2}\sigma_B \Big) \leqslant e^{-c_V \sigma_B}$，从而

$$\mathbb{P}\Big(\frac{1}{2}\sigma_B \leqslant V \leqslant 2\sigma_B \Big) \geqslant 1 - 2e^{-c_V \sigma_B}. \tag{10.20}$$

步骤 4：利用 $B \overset{d}{=} \sqrt{V}\, Z$ 和（10.20），我们将其与高斯尾部进行比较。对于上尾，对于任意 $t >\ 0,$

$$\mathbb{P}(|B| \geqslant t) = \mathbb{E}\big[\mathbb{P}(|Z| \geqslant t/\sqrt{V} \mid V) \big] \leqslant \mathbb{P}(V > 2\sigma_B) + \mathbb{P}\big(|Z| \geqslant t/\sqrt{2\sigma_B} \big).$$

取 $t := C_B \sqrt{\sigma_B D}$。于是 $t/\sqrt{2\sigma_B} = C_B \sqrt{D/2},$，而高斯尾界 $\mathbb{P}(|Z| \geqslant x) \leqslant 2e^{-x^2/2}$ 给出

$$\mathbb{P}(|B| \geqslant C_B \sqrt{\sigma_B D}) \leqslant e^{-c_V \sigma_B} + 2\exp\Big(-\frac{C_B^2}{4} D \Big) \leqslant e^{-D}$$

对于所有 D 在取 C_B 足够大、c 足够小之后。类似地，对于下尾

$$\mathbb{P}(|B| \geqslant t) \geqslant \mathbb{P}\big(V \geqslant \tfrac{1}{2}\sigma_B \big)\, \mathbb{P}\big(|Z| \geqslant t/\sqrt{\sigma_B/2} \big) \geqslant \frac{9}{10}\, \mathbb{P}\big(|Z| \geqslant \sqrt{2}\, t/\sqrt{\sigma_B} \big).$$

取 $t := c_B \sqrt{\sigma_B D}$。于是 $\sqrt{2}\, t/\sqrt{\sigma_B} = \sqrt{2}\, c_B \sqrt{D}$。根据标准高斯下尾，对所有 $x \geqslant 1$，$\mathbb{P}(|Z| \geqslant x) \geqslant (c_0/x) e^{-x^2/2}$。因此，当 $D \geqslant 1$ 和 c_B 足够小时，

$$\mathbb{P}(|B| \geqslant c_B \sqrt{\sigma_B D}) \geqslant \frac{c_1}{\sqrt{D}} \exp\Big(-c_2 D \Big) \geqslant e^{-D},$$

再次调整 常数。因此，在 $1 \leqslant D \leqslant c\sigma_B$ 的情形下，我们有常数 $0 < c_B < C_B$，使得

$$\mathbb{P}\big(|B| \geqslant c_B \sqrt{\sigma_B D} \big) \geqslant e^{-D}, \qquad \mathbb{P}\big(|B| \geqslant C_B \sqrt{\sigma_B D} \big) \leqslant e^{-D}.$$

步骤 5：现在我们准备好证明这两个界。固定 D，使得 $1 \leqslant D \leqslant c\min\{n,\ \sigma_B\}$。

上界：设 $a := C_A \sqrt{nD}$ 和 $b := C_B \sqrt{\sigma_B D}$。那么

$$\mathbb{P}(|S| \geqslant ab) \leqslant \mathbb{P}(|A| \geqslant a) + \mathbb{P}(|B| \geqslant b) \leqslant e^{-D} + e^{-D} \leqslant 2e^{-D},$$

因此（将因子 2 吸收到常数中），我们得到

$$q(D) \leqslant ab = C_A C_B \sqrt{n\sigma_B} D = C\sigma_S D.$$

下界：为了得到 $q(D)$ 的下界，令 $a := c_A\sqrt{nD}$ 且 $b := c_B\sqrt{\sigma_B D}$。由独立性，

$$\mathbb{P}(|S| \geqslant ab) \geqslant \mathbb{P}(|A| \geqslant a)\,\mathbb{P}(|B| \geqslant b) \geqslant e^{-D} \cdot e^{-D} = e^{-2D}.$$

现在将构造中的 D 替换为 $D/2$（其仍处于相同的范围内，至多相差常数因子），得到

$$\mathbb{P}\big(|S| \geqslant c\sqrt{n\sigma_B} D\big) \geqslant e^{-D}.$$

因此，$q(D) \geqslant c\sqrt{n\sigma_B}\, D = c\sigma_S D.$。结合这两个界可得，对于所有 $1 \leqslant D \leqslant c$，当 $\min\{n,\ \sigma_B\}$ 时，$c\,\sigma_S D \leqslant q(D) \leqslant C\,\sigma_S D$，如所声称。

10.3.5 引理 10.16 的证明

这个引理的证明相当技术性。我们将其拆解为若干个引理，并逐一陈述和证明，从而推导出所需结果。

我们首先为 χ^2 分布证明一个局部中心极限定理（中心极限定理）。

引理 10.17（关于 χ^2_u 在 $\sqrt{u}$ 尺度下的密度中心极限定理）。设 $u \in \mathbb{N}$ 满足 $u = \omega(1)$，并考虑 g_u 为某个 χ^2_u 分布的密度。假设 $D_n = o(u^{1/6})$。那么，当 $n \to \infty$ 时，对于 $|x| \leqslant D_n$，上述结论一致成立

$$\sup_{|x| \leqslant D_n} \Big|\sqrt{2u}\, g_u\big(u + x\sqrt{2u}\big) - \phi(x)\Big| \to 0, \tag{10.21}$$

$$\sup_{|x| \leqslant D_n} \Big|(2u)\, g_u'\big(u + x\sqrt{2u}\big) - \phi'(x)\Big| \to 0. \tag{10.22}$$

证明。写作 $u = 2\nu$，于是 $\nu = u/2 = \omega(1)$。对于 $t > 0$，χ^2_u 密度为

$$g_u(t) = \frac{1}{2^\nu \Gamma(\nu)} t^{\nu-1} e^{-t/2}.$$

固定 $|x| \leqslant D_n$ 并令 $t = u + x\sqrt{2u} = 2\nu + 2x\sqrt{\nu} = 2\nu\Big(1 + \frac{x}{\sqrt{\nu}}\Big)$。考虑对数密度

$$\log g_u(t) = -(\nu \log 2 + \log \Gamma(\nu)) + (\nu - 1)\log t - \frac{t}{2}.$$

由斯特林公式，$\Gamma(\nu) = \sqrt{2\pi}\,\nu^{\nu-\frac{1}{2}} e^{-\nu}\big(1 + o(1)\big)$，$\nu \to \infty$，这意味着 $\log \Gamma(\nu) = \Big(\nu - \frac{1}{2}\Big)\log \nu - \nu + \frac{1}{2}\log(2\pi) + o(1)$，当 $\nu \to \infty$ 时上述结论一致成立。此外，$\log t = \log(2\nu) + \log\Big(1 + \frac{x}{\sqrt{\nu}}\Big)$，并且对于 $|x| \leqslant D_n$ 一致成立，

$$\log\Big(1 + \frac{x}{\sqrt{\nu}}\Big) = \frac{x}{\sqrt{\nu}} - \frac{x^2}{2\nu} + O\Big(\frac{|x|^3}{\nu^{3/2}}\Big).$$

由于 $D_n = o(u^{1/6})$ 和 $u = 2\nu$，我们有

$$\sup_{|x| \leqslant D_n} \frac{|x|^3}{\nu^{3/2}} = o\Big(\frac{1}{\nu}\Big) \qquad \text{and in particular} \qquad \sup_{|x| \leqslant D_n} \frac{|x|^3}{\sqrt{\nu}} \to 0.$$

现在代入对数 $g_u(t)$，并注意到 $\frac{t}{2} = \nu + x\sqrt{\nu}$，得到

$$\begin{aligned}
\log g_u(t) &= -\nu \log 2 - \Big[\Big(\nu - \tfrac{1}{2}\Big)\log \nu - \nu + \tfrac{1}{2}\log(2\pi)\Big] \\
&\quad + (\nu - 1)\Big[\log(2\nu) + \frac{x}{\sqrt{\nu}} - \frac{x^2}{2\nu} + O\Big(\frac{|x|^3}{\nu^{3/2}}\Big)\Big] - \nu - x\sqrt{\nu} + o(1) \\
&= -\log 2 - \tfrac{1}{2}\log \nu - \tfrac{1}{2}\log(2\pi) + (\nu - 1)\frac{x}{\sqrt{\nu}} - x\sqrt{\nu} - (\nu - 1)\frac{x^2}{2\nu} \\
&\quad + O\Big(\frac{|x|^3}{\sqrt{\nu}}\Big) + O\Big(\frac{1 + x^2}{\nu}\Big) + o(1) \\
&= -\log 2 - \tfrac{1}{2}\log \nu - \tfrac{1}{2}\log(2\pi) - \frac{x^2}{2} + r_u(x),
\end{aligned}$$

其中

$$r_u(x) = -\frac{x}{\sqrt{\nu}} + O\Big(\frac{1+x^2}{\nu}\Big) + O\Big(\frac{|x|^3}{\sqrt{\nu}}\Big) + o(1).$$

现在加入归一化因子 $\sqrt{2u} = \sqrt{4\nu}$:

$$\begin{aligned}\log\Big(\sqrt{2u}\, g_u(t)\Big) &= \log g_u(t) + \tfrac{1}{2}\log(4\nu)\\ &= \Big(-\log 2 - \tfrac{1}{2}\log\nu - \tfrac{1}{2}\log(2\pi)\Big) + \Big(\log 2 + \tfrac{1}{2}\log\nu\Big) - \frac{x^2}{2} + r_u(x)\\ &= -\tfrac{1}{2}\log(2\pi) - \frac{x^2}{2} + r_u(x) = \log\phi(x) + r_u(x).\end{aligned}$$

因此， $\sqrt{2u}\, g_u(u + x\sqrt{2u}) = \phi(x)e^{r_u(x)}$. 于是,

$$\big|\sqrt{2u}\, g_u(u + x\sqrt{2u}) - \phi(x)\big| = \phi(x)\,|e^{r_u(x)} - 1| \leqslant \phi(0)\,|e^{r_u(x)} - 1|,$$

并对 $\sup_{|x|\leqslant D_n}$ 取上确界，即可证明第一个结论。

对于导数结论，定义

$$h_u(x) := \sqrt{2u}\, g_u(u + x\sqrt{2u}) = \phi(x)e^{r_u(x)}.$$

我们对 g_u 求导，得到

$$g_u'(t) = g_u(t)\left(\frac{\nu-1}{t} - \frac{1}{2}\right).$$

根据链式法则， $h_u'(x) = (2u)\, g_u'(u + x\sqrt{2u})$. 因此

$$h_u'(x) = h_u(x)\, A_u(x), \qquad A_u(x) := \sqrt{2u}\left(\frac{\nu-1}{t} - \frac{1}{2}\right).$$

利用 $u = 2\nu$ 和 $t = 2\nu(1+y)$,

$$A_u(x) = 2\sqrt{\nu}\left(\frac{\nu-1}{2\nu(1+y)} - \frac{1}{2}\right) = \sqrt{\nu}\left(\frac{1-1/\nu}{1+y} - 1\right) = -\frac{x + \nu^{-1/2}}{1 + x/\sqrt{\nu}}.$$

因此

$$A_u(x) + x = \frac{x^2 - 1}{\sqrt{\nu}\,(1 + x/\sqrt{\nu})}.$$

由于 D_n对于所有充分大的 ν，$/\sqrt{\nu} \to 0,$成立；对于所有 $|x| \leqslant D_n$，$1 + x/\nu \geqslant \frac{1}{2},$成立,，因此

$$\sup_{|x|\leqslant D_n} |A_u(x) + x| \leqslant \frac{2(D_n^2 + 1)}{\sqrt{\nu}} \to 0.$$

现在回忆一下， $h_u(x) = \phi(x)e^{r_u(x)}$ 和 $\phi'(x) = -x\phi(x)$。因此

$$\begin{aligned}h_u'(x) - \phi'(x) &= h_u(x)A_u(x) + x\phi(x)\\ &= \phi(x)e^{r_u(x)}A_u(x) + x\phi(x).\end{aligned}$$

加上再减去 $x\phi(x)e^{r_u(x)}$ ，得到

$$h_u'(x) - \phi'(x) = \phi(x)e^{r_u(x)}(A_u(x) + x) + x\phi(x)\big(1 - e^{r_u(x)}\big).$$

因此

$$|h_u'(x) - \phi'(x)| \leqslant \phi(x)e^{|r_u(x)|}\,|A_u(x) + x| + |x|\phi(x)\,|1 - e^{r_u(x)}|.$$

对 $|x| \leqslant D_n$ 取上确界，我们得到

$$\sup_{|x|\leqslant D_n} |h_u'(x) - \phi'(x)| \leqslant \Big(\sup_{|x|\leqslant D_n} \phi(x)\Big)\Big(\sup_{|x|\leqslant D_n} e^{|r_u(x)|}\Big)\Big(\sup_{|x|\leqslant D_n} |A_u(x)+x|\Big) + \Big(\sup_{|x|\leqslant D_n} |x|\phi(x)\Big)\Big(\sup_{|x|\leqslant D_n} |1-e^{r_u(x)}|\Big).$$

现在 $\sup_{|x|\leqslant D_n} \phi(x) \leqslant \frac{1}{\sqrt{2\pi}}$. 此外， $|x|\phi(x) = \frac{|x|}{\sqrt{2\pi}} e^{-x^2/2}$. 对于 $x \geqslant 0$，函数 $x \mapsto xe^{-x^2/2}$ 的导数为 $\frac{d}{dx}\big(xe^{-x^2/2}\big) = e^{-x^2/2}(1-x^2),$ ，因此其最大值在 $x = 1$处取得。于是

$$\sup_{|x|\leqslant D_n} |x|\phi(x) \leqslant \frac{e^{-1/2}}{\sqrt{2\pi}}.$$

由于 $\sup_{|x|\leqslant D_n} |r_u(x)| \to 0$，我们有

$$\sup_{|x|\leqslant D_n} e^{|r_u(x)|} \to 1 \qquad \text{and} \qquad \sup_{|x|\leqslant D_n} |1-e^{r_u(x)}| \to 0.$$

此外，我们已经 证明了 $\sup_{|x|\leqslant D_n} |A_u(x)+x| \to 0.$。结合这些界，我们得出结论 t

$$\sup_{|x|\leqslant D_n} |h_u'(x) - \phi'(x)| \to 0.$$

由于 $h_u'(x) = (2u)\, g_u'(u + x\sqrt{2u}),$ 这就证明了

$$\sup_{|x|\leqslant D_n} \Big|(2u)\, g_u'(u+x\sqrt{2u}) - \phi'(x)\Big| \to 0.$$

这就证明了（10.22）。 □

设 G、$H \sim N(0、1)$ 相互独立且 $Y := GH$。对于整数 $u \geqslant 1$，令 Y_1、Y_2、.....为 Y 的独立同分布副本，并定义

$$B_u := \sum_{k=1}^{u} Y_k. \tag{10.23}$$

令 f_u 为 B_u 在 $u \geqslant 1$下的密度。现在我们证明 B_u的局部中心极限定理。回顾引理 9.5，$B_u = (X_u - Y_u)/2$ 其中 X_u， Y_u 独立同分布 χ^2_u。我们在此陈述该引理，其证明将在第 10.4节给出。

引理 10.18 （关于 f_u 和 f_u'的密度中心极限定理）。 假设 $D_n = o(u^{1/6})$。对于 $|x| \leqslant D_n$，一致地有

$$\sup_{|x|\leqslant D_n} \big|\sqrt{u} f_u(x\sqrt{u}) - \phi(x)\big| \to 0, \tag{10.24}$$

$$\sup_{|x|\leqslant D_n} \big|u f_u'(x\sqrt{u}) - \phi'(x)\big| \to 0. \tag{10.25}$$

我们还将使用以下关于 B_u的密度及其导数的界。该结果的证明也将在第 10.4节给出。

引理 10.19 （远离零点时关于 f_u 和 f_u' 的一致界）。 对于每个 $u \geqslant 1$， B_u 的密度 f_u 在 $\mathbb{R}\backslash\{0\}$ 上可微。此外，存在一个绝对常数 $C > 0$ ，使得对所有 $u \geqslant 1$ 和所有 $x \neq 0$，

$$f_u(x) \leqslant \frac{C}{|x|}, \qquad |f_u'(x)| \leqslant \frac{C}{x^2}.$$

现在我们可以继续讨论稀疏聚类模型中出现在重叠里的随机变量 B 的密度。

引理 10.20 （ B 条件化的精确密度）。 设 f_B 表示 B 在 $\mathbb{R}\backslash\{0\}$ 上的密度。那么对于所有$x \neq 0$,

$$f_B(x) = \sum_{u=1}^{p} \mathbb{P}(U = u) f_u(x), \qquad f'_B(x) = \sum_{u=1}^{p} \mathbb{P}(U = u) f'_u(x).$$

证明。 根据定义， $B \mid (U = u) \stackrel{d}{=} B_u$。对于任意博雷尔集 $E \subset \mathbb{R}\backslash\{0\}$,

$$\mathbb{P}(B \in E) = \sum_{u=1}^{p} \mathbb{P}(U = u)\mathbb{P}(B_u \in E),$$

因此，将等式写成积分形式，并利用概率密度函数的唯一性，可得 $f_B(x) = \sum_{u=1}^{p} \mathbb{P}(U = u) f_u(x)$ 对所有 $x \neq 0$均成立。由于求和项数有限，我们逐项对 $x \neq 0$ 求导，得到 $f'_B(x)$的公式。□

引理 10.21（ B 和 f'_B 在 $\sqrt{\sigma_B}$ 尺度下的密度中心极限定理）。固定 $0 < \varepsilon < 1$。假设 $D_n = o(\sigma_B^{1/6})$。对 $\varepsilon \leqslant |x| \leqslant D_n$，一致地成立

$$\sup_{\varepsilon \leqslant |x| \leqslant D_n} \left|\sqrt{\sigma_B} f_B(x\sqrt{\sigma_B}) - \phi(x)\right| \to 0, \tag{10.26}$$

$$\sup_{\varepsilon \leqslant |x| \leqslant D_n} \left|\sigma_B f'_B(x\sqrt{\sigma_B}) - \phi'(x)\right| \to 0. \tag{10.27}$$

证明。 令 $L =$对数 σ_B并令$\mathcal{I}$：$= \{u : |u - \sigma_B| \leqslant \sqrt{\sigma_B} L\}$。我们将引理 10.20中的混合分解为 $u \in \mathcal{I}$和 $u \notin \mathcal{I}$。

步骤1： 由二项分布的切尔诺夫不等式，$\mathbb{P}(U \notin \mathcal{I}) \leqslant 2e^{-cL^2} = o(1)$。由引理10.19，对每个 $u \geqslant 1$ 以及每个 $z \neq 0$,

$$f_u(z) \leqslant \frac{C}{|z|}, \qquad |f'_u(z)| \leqslant \frac{C}{z^2}.$$

因此，对于 $\varepsilon \leqslant |x| \leqslant D_n$,

$$|f_u(x\sqrt{\sigma_B})| \leqslant \frac{C}{\varepsilon\sqrt{\sigma_B}}, \qquad |f'_u(x\sqrt{\sigma_B})| \leqslant \frac{C}{\varepsilon^2\sigma_B}.$$

定义 C_ε：$= C/\varepsilon^2$. 然后利用引理 10.20 可得，对 $\varepsilon \leqslant |x| \leqslant D$一致成立 n,

$$\sqrt{\sigma_B} \cdot \sum_{u \notin \mathcal{I}} \mathbb{P}(U = u)\, |f_u(x\sqrt{\sigma_B})| \leqslant C_\varepsilon\, \mathbb{P}(U \notin \mathcal{I}) = o(1),$$

类似地

$$\sigma_B \cdot \sum_{u \notin \mathcal{I}} \mathbb{P}(U = u)\, |f'_u(x\sqrt{\sigma_B})| \leqslant C_\varepsilon\, \mathbb{P}(U \notin \mathcal{I}) = o(1).$$

步骤 2：固定 $\varepsilon \leqslant |x| \leqslant D_n$ 和 $u \in \mathcal{I}$。记

$$z_u(x) := x\sqrt{\sigma_B/u}.$$

由于 $u = \sigma_B + O(\sqrt{\sigma_B L})$，我们有 $z_u(x) = x + O(|x|L/\sqrt{\sigma_B})$, 对$\varepsilon \leqslant |x| \leqslant D_n$一致成立。因为$|x| \leqslant D_n$ 且 $D_n L/\sqrt{\sigma_B} = o(1)$, 我们有 $|z_u(x)| \leqslant 2D_n$ 对所有充分大的 $(n,\ p)$。此外，由于 $|x| \geqslant \varepsilon$，对所有充分大的$(n,\ p)$ 我们有 $|z_u(x)| \geqslant \varepsilon/2$。对10.18 一致地应用引理，令 $\varepsilon/2 \leqslant |z| \leqslant 2D_n$:

$$f_u(x\sqrt{\sigma_B}) = f_u(z_u(x)\sqrt{u}) = \frac{1}{\sqrt{u}}\phi(z_u(x)) + o\Big(\frac{1}{\sqrt{u}}\Big),$$

对 $\varepsilon \leqslant |x| \leqslant D_n$ 和 $u \in \mathcal{I}$均匀成立。接下来，由于 $\sup_{\varepsilon \leqslant |x| \leqslant D_n} \sup_{u \in \mathcal{I}} |z_u(x) - x| = o(1)$ 和 ϕ 是 $O(1)$利普希茨的，我们有

$$\sup_{\varepsilon \leqslant |x| \leqslant D_n} \sup_{u \in \mathcal{I}} |\phi(z_u(x)) - \phi(x)| = o(1).$$

此外，对 $u \in \mathcal{I}$均匀成立，我们有 $1/\sqrt{u} = 1/\sqrt{\sigma_B} + O(L/\sigma_B)$。结合这两点可得

$$f_u(x\sqrt{\sigma_B}) = \frac{1}{\sqrt{\sigma_B}}\phi(x) + o\Big(\frac{1}{\sqrt{\sigma_B}}\Big),$$

在 $u \in \mathcal{I}$ 和 $\varepsilon \leqslant |x| \leqslant D_n$上均匀成立。用权重 $\mathbf{P}(U = u)$ 在 $u \in \mathcal{I}$ 上取平均，并应用步骤1（注意， U 的几乎所有质量都位于 $\mathcal{I}$ 内），得到（10.26）。关于导数的结论（10.27）同理可得，使用引理 10.18 ，特别是（10.25）：

$$f_u'(x\sqrt{\sigma_B}) = f_u'(z_u(x)\sqrt{u}) = \frac{1}{u}\phi'(z_u(x)) + o\Big(\frac{1}{u}\Big),$$

然后再次利用 ϕ' 是 $O(1)$ 利普希茨这一事实，得到 $1/u = 1/\sigma_B + O(L/\sigma_B^{3/2})$ 和 $\phi'(z_u(x)) = \phi'(x) + o(1)$。 □

引理 10.22 （ A 的局部中心极限定理及尾部）。 设 $T := \sqrt{D_n}$ 对数$(n \vee p)$。对满足 $|a| \leqslant \sqrt{n}T$ 的整数 $a \equiv n$ （模 2），以下结论均匀成立，

$$\mathbb{P}(A = a) = \frac{2}{\sqrt{n}}\phi\Big(\frac{a}{\sqrt{n}}\Big)\Big(1 + o(1)\Big).$$

此外 $\mathbf{P}(|A| \geqslant \sqrt{n}T) \leqslant 2e^{-T^2/2}$。

证明。 第一部分是引理的标准应用10.1。对于尾部，独立同分布的 ± 1之和的 Hoeffding不等式给出

$$\mathbb{P}(|A| \geqslant \sqrt{n}T) \leqslant 2e^{-T^2/2}.$$

□

引理 10.23 (混合密度和导数)。 令 f_B 为 B 的密度。那么对于每个 $t \neq 0$,

$$f_S(t) = \sum_{a \in \{-n, -n+2, \ldots, n\} \setminus \{0\}} \mathbb{P}(A = a)\frac{1}{|a|} f_B(t/a),$$

且 f_S 在 $t \neq 0$ 上可微，满足

$$f_S'(t) = \sum_{a \neq 0} \mathbb{P}(A = a)\frac{1}{a|a|} f_B'(t/a).$$

证明。 由于 $S = A \cdot B$，我们通过对 A 进行条件化来证明这一点。根据全概率公式以及 A 和 B 的独立性，对于任意 $x \in \mathbb{R}$,

$$\mathbb{P}(S \leqslant x) = \sum_{a \in \{-n, -n+2, \ldots, n\}} \mathbb{P}(A = a)\,\mathbb{P}(aB \leqslant x).$$

该和是有限的，因此我们可以在任意 $t \neq 0$ 处逐项求导，得到

$$f_S(t) = \frac{d}{dt}\,\mathbb{P}(S \leqslant t) = \sum_{a \neq 0} \mathbb{P}(A = a)\,\frac{1}{|a|}\,f_B(t/a),$$

这就是 $f_S(t)$ 在 $t \neq 0$ 上的所述公式。对于 f_S'，只需再对 $t \neq 0$ 逐项求一次导数即可直接得到结果。□

令 $G, H \sim N(0,1)$ 相互独立，并令 $W := GH$。其密度为

$$f_\infty(x) = \int_{\mathbb{R}} \frac{1}{|y|}\phi(y)\phi(x/y)dy, \qquad x \neq 0. \tag{10.28}$$

引理 10.24 (密度局部极限定理（*LLT*） （适用于 S 在 $t = D_n\sigma_S$处）。 假设 $D_n = o(\sigma_B^{1/6})$。设$t := D_n\sigma_S = D_n\sqrt{n\sigma_B}$，其中随机变量 B， S 定义见 *10.3.1*。那么，

$$\sigma_S f_S(t) = f_\infty(D_n)\big(1 + o(1)\big), \qquad \sigma_S^2 f_S'(t) = f_\infty'(D_n)\big(1 + o(1)\big). \tag{10.29}$$

证明。 固定 $t = D_n\sigma_S = D_n\sqrt{n\sigma_B}$，并令 $T = \sqrt{D_n}$对数$(n \vee p)$ 如引理 10.22中所述。我们将反复使用初等下界（对所有 $x \geqslant 1$均有效）

$$f_\infty(x) = \int_{\mathbb{R}} \frac{1}{|u|}\phi(u)\phi(x/u)\,du \;\geqslant\; \int_{\sqrt{x}}^{\sqrt{x}+1} \frac{1}{u}\phi(u)\phi(x/u)\,du \geqslant \frac{c}{\sqrt{x}}\,e^{-Cx}, \tag{10.30}$$

对某些绝对常数 c，$C > 0$（因为在 $[\sqrt{x},\ \sqrt{x+1}]$ 上有 $u = \Theta(\sqrt{x})$, $x/u = \sqrt{x} \mp O(1),$，因此$\phi(u)\phi(x/u) = \Theta(e^{-x})$）。

最后请注意，对于每个固定的 $x \neq 0$，都可以在积分号下对（10.28）求导（因为当$u \to 0$ 时，因子 $\phi(x/u)$ 超指数地衰减；当 $|u| \to \infty$ 时，因子 $\phi(u)$ 衰减），从而得到$x \neq 0$,

$$f'_\infty(x) = \int_{\mathbb{R}} \frac{1}{u|u|}\phi(u)\phi'(x/u)\,du. \tag{10.31}$$

此外，将（10.31）对 $u \in [\sqrt{x}$ 积分，$\sqrt{x+1}]$ 注意到 $\phi'(x/u) = \Theta(-\sqrt{x}\,\phi(\sqrt{x}))$，然后再次使用（10.30）

$$f'_\infty(x) \geqslant C_1 \int_{\sqrt{x}}^{\sqrt{x}+1} \frac{1}{u}\phi(u)\phi(x/u)\,du \geqslant \frac{c'}{\sqrt{x}}\,e^{-Cx}, \tag{10.32}$$

对某些绝对常数 C_1，$c' > 0$。

步骤1： 首先，我们要证明尾部贡献 $|a| > \sqrt{n}\,T$ 非常小。为此，我们将在 t/a处应用引理 10.22,

$$\mathbb{P}(|A| \geqslant \sqrt{n}\,T) \leqslant 2e^{-T^2/2} = 2\exp\Big(-\frac{1}{2}D_n \log^2(n \vee p)\Big).$$

对于 $|a| > \sqrt{n}\,T$，定义

$$x_a := \frac{t/a}{\sqrt{\sigma_B}} = \frac{D_n\sqrt{n\sigma_B}/a}{\sqrt{\sigma_B}} = \frac{D_n}{a/\sqrt{n}}, \qquad \text{so} \qquad |x_a| \leqslant \frac{D_n}{T} = \frac{\sqrt{D_n}}{\log(n \vee p)}.$$

特别地， $|x_a| \leqslant D_n$，因此引理 10.21 对 的选择均适用，并得到

$$|f_B(t/a)| \leqslant \frac{C}{\sqrt{\sigma_B}}, \qquad |f'_B(t/a)| \leqslant \frac{C}{\sigma_B}\,|x_a| \leqslant \frac{C}{\sigma_B}\cdot\frac{D_n}{T},$$

对于某个绝对常数 $C > 0$，利用 $|\phi(x)| \leqslant (2\pi)^{-1/2}$ 以及 $|\phi'(x)| = |x|\phi(x) \leqslant C|x|$。因此，利用引理 10.23 以及 $|a| > \sqrt{n}\,T \Rightarrow |a|^{-1} \leqslant (\sqrt{n}\,T)^{-1}$,

$$\begin{aligned}\sigma_S \sum_{|a|>\sqrt{n}\,T} \mathbb{P}(A = a)\frac{1}{|a|}\,|f_B(t/a)| &\leqslant \sigma_S \cdot \mathbb{P}(|A| > \sqrt{n}\,T)\cdot\frac{1}{\sqrt{n}\,T}\cdot\frac{C}{\sqrt{\sigma_B}} \\ &= \frac{C}{T}\,\mathbb{P}(|A| > \sqrt{n}\,T) = o\Big(\frac{e^{-D_n}}{\sqrt{D_n}}\Big) = o\big(f_\infty(D_n)\big),\end{aligned} \tag{10.33}$$

其中我们使用了 $\sigma_S = \sqrt{n\sigma_B}$和（10.30）。类似地，由于 $|a| > \sqrt{n}T \Rightarrow |a|^{-2} \leqslant (nT^2)^{-1}$,

$$\begin{aligned}\sigma_S^2 \sum_{|a|>\sqrt{n}\,T} \mathbb{P}(A = a)\frac{1}{|a|^2}\,|f'_B(t/a)| &\leqslant \sigma_S^2 \cdot \mathbb{P}(|A| > \sqrt{n}\,T)\cdot\frac{1}{nT^2}\cdot\frac{C}{\sigma_B}\cdot\frac{D_n}{T} \\ &= C\,\frac{D_n}{T^3}\,\mathbb{P}(|A| > \sqrt{n}\,T) = o\Big(\frac{e^{-D_n}}{\sqrt{D_n}}\Big) = o\big(|f'_\infty(D_n)|\big),\end{aligned} \tag{10.34}$$

其中最后一步使用了 (10.32)。因此，从现在起，我们将求和范围限制为 $|a| \leqslant \sqrt{n}\,T$。

步骤 2：令 $y :\, = a/\sqrt{n}$ 于是 $|y| \leqslant T$ 并且 $|a| = \sqrt{n}|y|$。然后

$$\frac{t}{a} = \frac{D_n\sqrt{n\sigma_B}}{\sqrt{n}y} = \frac{D_n\sqrt{\sigma_B}}{y}.$$

引理 10.22 表明 $\mathbf{P}(A=a)=(2/\sqrt{n})\phi(y)(1+o(1))$ 在 $|y|\leqslant T$ 上均匀成立。令

$$\mathcal{A}:=\Big\{\alpha\in\mathbb{Z}:\tfrac{1}{2}\sqrt{nD_n}\leqslant|\alpha|\leqslant 2\sqrt{nD_n}\Big\}.$$

我们将进一步把求和截断至 $\alpha\in\mathcal{A}$。我们声称，即使经过截断 $|\alpha|\leqslant\sqrt{nT}$，$\alpha\notin\mathcal{A}$ 对 $\sigma_S f_S(t)$ 的贡献仍然是 $o(f_\infty(D_n))$。事实上，再次使用符号说明 $y=a/\sqrt{n}$，则

$$g_{D_n}(y):=\frac{1}{|y|}\phi(y)\phi(D_n/y).$$

对于 $|y|\leqslant\frac{1}{2}\sqrt{D_n}$，我们有 $(D_n/y)^2\geqslant 4D_n$，因此

$$g_{D_n}(y)\leqslant\frac{C}{|y|}\exp\Big(-\frac{1}{2}\cdot\frac{D_n^2}{y^2}\Big)\leqslant\frac{C}{|y|}e^{-2D_n},\tag{10.35}$$

并对 $0<|y|\leqslant\frac{1}{2}\sqrt{D_n}$ 积分，得到 $\int_{0<|y|\leqslant\frac{1}{2}\sqrt{D_n}}g_{D_n}(y)\,dy\leqslant Ce^{-2D_n}$ 对数 D_n。类似地，对于 $|y|\geqslant 2\sqrt{D_n}$，我们有 $y^2\geqslant 4D_n$，因此 $g_{D_n}(y)\leqslant Ce^{-2D_n}/|y|$，从而 $\int_{|y|\geqslant 2\sqrt{D_n}}g_{D_n}(y)\,dy\leqslant Ce^{-2D_n}$。与（10.30）比较可知，两侧尾部均为 $o(f_\infty(D_n))$。固定 $\alpha\in\mathcal{A}$ 及其对应的点 $y=\alpha/\sqrt{n}$。由引理 10.22（利用 $\mathcal{A}\subseteq\{|\alpha|\leqslant\sqrt{nT}\}$ 对所有充分大的 n），

$$\mathbb{P}(A=a)=\frac{2}{\sqrt{n}}\phi(y)\big(1+o(1)\big),\tag{10.36}$$

接下来，注意对于 $\alpha\in\mathcal{A}$，我们有

$$\frac{t}{a}=\frac{D_n\sqrt{n\sigma_B}}{\sqrt{n}\,y}=\frac{D_n\sqrt{\sigma_B}}{y},\qquad\text{so}\qquad\frac{t/a}{\sqrt{\sigma_B}}=\frac{D_n}{y}\in\Big[\tfrac{1}{2}\sqrt{D_n},\,2\sqrt{D_n}\Big].$$

特别地，该比值至多为 $2\sqrt{D_n}=o(\sigma_B^{1/6})$，因为 $D_n=o(\sigma_B^{1/6})$。因此，我们可以应用引理10.21，并对导数进行类似处理，得到：

$$f_B\Big(\frac{D_n\sqrt{\sigma_B}}{y}\Big)=\frac{1}{\sqrt{\sigma_B}}\phi(D_n/y)\big(1+o(1)\big),\qquad f_B'\Big(\frac{D_n\sqrt{\sigma_B}}{y}\Big)=\frac{1}{\sigma_B}\phi'(D_n/y)\big(1+o(1)\big).\tag{10.37}$$

将（10.36）和（10.37）代入引理 10.23，并利用 $|a|=\sqrt{n}\,|y|$，得到

$$f_S(t)=\frac{(1+o(1))}{\sigma_S}\cdot\sum_{\substack{0<|a|\leqslant\sqrt{n}T\\ a\in\mathcal{A}}}\Big(\frac{2}{\sqrt{n}}\Big)\frac{1}{|y|}\phi(y)\phi(D_n/y).$$

该和是在 $x=D_n$ 处对（10.28）求值所得到的黎曼和，因而得到（10.29）。

步骤 3：对于导数，我们使用引理 10.23 对 f_S' 重复这一过程。

我们首先证明，$\alpha\notin\mathcal{A}$ 的贡献对于导数同样可以忽略。对于 $y=a/\sqrt{n}$，$0<|y|\leqslant\frac{1}{2}\sqrt{D_n}$，利用（10.35）中的界，

$$\frac{D_n}{|y|^2}g_{D_n}(y)\leqslant C\frac{D_n}{|y|^3}\exp\Big(-\frac{D_n^2}{2y^2}\Big).$$

因此，通过变量替换 $z=D_n/|y|$，

$$\int_{0<|y|\leqslant\frac{1}{2}\sqrt{D_n}}\frac{D_n}{|y|^2}g_{D_n}(y)\,dy\leqslant C\int_{2\sqrt{D_n}}^{\infty}\frac{z}{D_n}\phi(z)\,dz\leqslant C\frac{e^{-2D_n}}{D_n}.$$

类似地，对于 $|y|\geqslant 2\sqrt{D_n}$，我们有

$$\frac{D_n}{|y|^2}g_{D_n}(y)\leqslant C\frac{D_n}{|y|^3}e^{-y^2/2},$$

因此

$$\int_{|y|\geqslant 2\sqrt{D_n}} \frac{D_n}{|y|^2} g_{D_n}(y)\,dy \leqslant CD_n \int_{|y|\geqslant 2\sqrt{D_n}} \frac{e^{-y^2/2}}{|y|^3}\,dy \leqslant C\frac{e^{-2D_n}}{D_n}.$$

与 (10.32) 相比，两端的尾部都是 $o(|f'_\infty(D_n)|)$。

现在，对于 $\alpha \in \mathcal{A}$，由 (10.36) 和 (10.37)，

$$\mathbb{P}(A = a) = \frac{2}{\sqrt{n}}\phi(y)\big(1+o(1)\big), \qquad f'_B\Big(\frac{D_n\sqrt{\sigma_B}}{y}\Big) = \frac{1}{\sigma_B}\phi'(D_n/y)\big(1+o(1)\big),$$

在 $\alpha \in \mathcal{A}$ 上均匀分布。将其代入关于 10.23 的引理，其中 f'_S，并利用 $y = \alpha/\sqrt{n}$，得到

$$f'_S(t) = \frac{(1+o(1))}{\sigma_S^2} \cdot \sum_{\substack{0<|a|\leqslant\sqrt{n}T \\ a\in\mathcal{A}}} \Big(\frac{2}{\sqrt{n}}\Big)\frac{1}{y|y|}\phi(y)\phi'(D_n/y).$$

该和是对 (10.31) 在 $x = D_n$ 处求值得到的黎曼和；根据刚刚证明的尾部界，$y \notin [\frac{1}{2}\sqrt{D_n},\ 2\sqrt{D_n}]$ 的贡献可以忽略。因此，

$$f'_S(t) = \frac{(1+o(1))}{\sigma_S^2} f'_\infty(D_n),$$

这就是所需的结论。 □

现在我们可以证明主要引理了。

引理 *10.16* 的证明。 根据引理 10.24，

$$f_S(D_n\sigma_S) = \frac{1}{\sigma_S}f_\infty(D_n)(1+o(1)), \qquad f'_S(D_n\sigma_S) = \frac{1}{\sigma_S^2}f'_\infty(D_n)(1+o(1)).$$

因此

$$\frac{d}{ds}\log f_S(s)\Big|_{s=D_n\sigma_S} = \frac{f'_S(D_n\sigma_S)}{f_S(D_n\sigma_S)} = \frac{1}{\sigma_S}\frac{f'_\infty(D_n)}{f_\infty(D_n)} + o\Big(\frac{1}{\sigma_S}\Big) = \frac{1}{\sigma_S}(\log f_\infty)'(D_n) + o\Big(\frac{1}{\sigma_S}\Big).$$

最后，根据（9.10），我们有 $f_\infty(x) = \frac{1}{\pi}K_0(|x|)$，并且对于 $x > 0$，

$$(\log f_\infty)'(x) = -\frac{K_1(x)}{K_0(x)}.$$

此外，根据引理9.2，

$$\frac{K_1(x)}{K_0(x)} \to 1 \qquad \text{as } x \to \infty.$$

因此，当 $x \to \infty$时，(对数 $f_\infty)'(x) \to -1$，所以将 σ_S重新缩放后，最高阶项为 $-1/\sigma_S$。 □

10.4 辅助引理

本节给出引理9.5、10.1、10.18和10.19的证明。

引理*9.5*的证明。 设

$$U = \frac{G+H}{\sqrt{2}}, \qquad V = \frac{G-H}{\sqrt{2}}.$$

那么 U，V是独立的 $N(0,\ I_d)$向量，并且

$$4\langle G, H\rangle = \|G+H\|^2 - \|G-H\|^2 = 2\|U\|^2 - 2\|V\|^2.$$

因此 $W=(X-Y)/2$, 其中 $X=\|U\|^2$ 和 $Y=\|V\|^2$ 是独立同分布的 χ^2_d。

接下来，对于单个坐标乘积 $Z=G_1H_1$,

$$\mathbb{E}[e^{itZ}\mid G_1]=\exp\left(-\frac{t^2G_1^2}{2}\right),$$

因此，对 $G_1\sim N(0、1)$ 取平均后，

$$\phi_Z(t):=\mathbb{E}[e^{itZ}]=(1+t^2)^{-1/2}.$$

由于 $W=\sum_{i=1}^d G_iH_i$ 是 d 个 Z 的独立副本之和，

$$\phi_W(t)=\mathbb{E}[e^{itW}]=(1+t^2)^{-d/2}.$$

现在使用标准的余弦变换恒等式

$$\int_0^\infty \frac{\cos(tx)}{(1+t^2)^\mu}\,dt=\frac{\sqrt{\pi}}{\Gamma(\mu)}\left(\frac{|x|}{2}\right)^{\mu-\frac12}K_{\mu-\frac12}(|x|),\qquad \mu>0,\ x\neq 0, \tag{10.38}$$

参见 [AS65，第 9]章。由于 ϕ_W 是偶函数，

$$f_d(x)=\frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}(1+t^2)^{-d/2}\,dt=\frac{1}{\pi}\int_0^\infty\frac{\cos(tx)}{(1+t^2)^{d/2}}\,dt.$$

将 (10.38) 应用于 $\mu=d/2$ 可得 (9.10)。由 (9.6) 可知，在 0 处的连续性结论成立。 □

引理的证明 *10.1*。对于 $s\equiv n\ (\text{模}\ 2)$，我们有

$$\mathbb{P}(S_n=s)=2^{-n}\binom{n}{j},\qquad j=\frac{n+s}{2},\quad p=\frac{j}{n}. \tag{10.39}$$

假设 $|s|\leqslant n/2,$，因此 $p\in[1/4,3/4]$ 且 $j,\ n-j\geqslant n/4$。利用带余项的斯特林公式，

$$\log(m!)=m\log m-m+\tfrac12\log(2\pi m)+O(1/m),$$

对 $p\in[1/4,3/4]$,一致地成立

$$\log\binom{n}{j}=nH(p)-\tfrac12\log\big(2\pi np(1-p)\big)+O(1/n), \tag{10.40}$$

其中 $H(p)=-p$ 对数 $p-(1-p)$对数$(1-p)$。结合（10.39）-（10.40）可得如下形式

$$\mathbb{P}(S_n=s)=\frac{1}{\sqrt{2\pi np(1-p)}}\exp\big(n(H(p)-\log 2)\big)\exp\big(O(1/n)\big),\qquad (|s|\leqslant n/2). \tag{10.41}$$

将 $p=\frac12+x$ 写作 $x:=\frac{s}{2n}$—(因此当 $|s|\leqslant n/2$时 $|x|\leqslant 1/4$）。将 H 在 $1/2$ 处展开泰勒级数，得到

$$n(H(p)-\log 2)=-\frac{s^2}{2n}+O\Big(\frac{|s|^3}{n^2}\Big),\qquad (|s|\leqslant n/2), \tag{10.42}$$

在该范围内关于 s 一致成立。特别地，这也意味着

$$n(H(p)-\log 2)\geqslant-\frac{s^2}{2n}-C\frac{|s|^3}{n^2},\qquad (|s|\leqslant n/2), \tag{10.43}$$

其中 $C>0$ 为绝对常数。

第一个断言的证明：假设 $|s|\leqslant\sqrt{n}\,T$满足 $T^3=o(\sqrt{n})$。那么 $|s|=o(n^{2/3})$ 特别是 $|s|\leqslant n/2$对所有充分大的 n，因此（10.41）适用。此外 $p(1-p)=\frac14+O(s^2/n^2)$, 因此

$$\frac{1}{\sqrt{2\pi np(1-p)}}=\frac{2}{\sqrt{2\pi n}}\exp\Big(O\Big(\frac{s^2}{n^2}\Big)\Big)=\frac{2}{\sqrt{2\pi n}}\big(1+o(1)\big),$$

在 $|s| \leqslant \sqrt{n}\,T$上一致成立，因为 $s^2/n^2 \leqslant T^2/n = o(1)$。由（10.42），

$$\frac{|s|^3}{n^2} \leqslant \frac{(\sqrt{n}\,T)^3}{n^2} = \frac{T^3}{\sqrt{n}} = o(1),$$

因此，(10.41) 中的指数余项在同一窗口上 $1 + o(1)$ 均匀成立。因此，

$$\mathbb{P}(S_n = s) = \frac{2}{\sqrt{2\pi n}} e^{-s^2/(2n)}\big(1 + o(1)\big) = \frac{2}{\sqrt{n}}\phi\Big(\frac{s}{\sqrt{n}}\Big)\big(1 + o(1)\big),$$

对 $|s| \leqslant \sqrt{n}\,T$均匀成立，从而证明了 (10.1)。

第二个结论的证明：取 n 充分大，并取 $0 \leqslant s \leqslant n/2$。于是 $p \in [1/4, 3/4]$，且 (10.41) 成立。利用粗略界 $p(1-p) \leqslant 1/4$ 可得

$$\frac{1}{\sqrt{2\pi n p(1-p)}} \geqslant \frac{2}{\sqrt{2\pi n}}.$$

将此与单侧熵界 (10.43) 以及因子 $\exp(O(1/n)) \geqslant 1/2$ 结合，可得

$$\mathbb{P}(S_n = s) \geqslant \frac{c}{\sqrt{n}} \exp\Big(-\frac{s^2}{2n} - C\frac{s^3}{n^2}\Big),$$

其中 c为绝对常数，$C > 0$ 且对所有 $n \geqslant n_\star$均成立，这就是 (10.2)。 □

在证明之前，我们先陈述两个辅助引理。它们的证明推迟到第10.4.1节。

引理 10.25（高斯乘积积分）。设 $\phi(z) := (2\pi)^{-1/2} e^{-z^2/2}$ 为标准正态密度。则对于任意 $x \in \mathbb{R}$,

$$\int_{\mathbb{R}} \phi(z)\,\phi(z - \sqrt{2}\,x)\,dz = \frac{1}{\sqrt{2}}\,\phi(x).$$

引理 10.26（χ^2_u 密度及其导数的上确界界）。设 g_u 为一个 χ^2_u 随机变量的密度：

$$g_u(t) = \frac{1}{2^{u/2}\Gamma(u/2)}\, t^{u/2-1} e^{-t/2}, \qquad t > 0.$$

则存在一个绝对常数 $C < \infty$，使得对所有 u,

$$\sup_{t>0} g_u(t) \;\leqslant\; \frac{C}{\sqrt{u}}, \qquad \sup_{t>0} |g_u'(t)| \;\leqslant\; \frac{C}{u}.$$

引理 *10.18*的证明。根据引理 9.5，$B_u = (X_u - Y_u)/2$ 其中 X_u，Y_u 独立同分布，χ^2_u 且密度为 g_u。因此，对于任意 $b \in \mathbb{R}$,

$$f_u(b) = 2\int_0^\infty g_u(t)\, g_u(t - 2b)\, \mathbf{1}\{t - 2b > 0\}\, dt. \tag{10.44}$$

固定 $|x| \leqslant D_n$ 并令 $b = x\sqrt{u}$。记 $u = 2\nu$。设 $t = u + z\sqrt{2u}$，$dt = \sqrt{2u}\,dz$. 约束 $t > 0$ 等价于 $z > -\sqrt{u/2}$。约束 $t - 2b > 0$ 等价于

$$u + z\sqrt{2u} - 2x\sqrt{u} > 0 \;\Leftrightarrow\; z > -\sqrt{u/2} + \sqrt{2}\,x.$$

下文中我们记$z_{u,\,x} := -\sqrt{u/2} + \sqrt{2}\,x$。由于 $|x| \leqslant D_n = o(u^{1/6})$，，我们有 $|x| = o(\sqrt{u})$，因此

$$\sup_{|x| \leqslant D_n} z_{u,x} = -\sqrt{u/2} + o(\sqrt{u}) \to -\infty.$$

因此，经过此变量替换，10.44变为

$$f_u(x\sqrt{u}) = 2\int_{z > z_{u,x}} g_u(u + z\sqrt{2u})\, g_u\big(u + (z - \sqrt{2}x)\sqrt{2u}\big)\,\sqrt{2u}\,dz. \tag{10.45}$$

固定 $M \geqslant 1$。将（10.45）中的积分拆分为 $|z| \leqslant M$ 和 $|z| > M$。

对于尾部 $|z| > M$，我们用 g_u 上的均匀上确界界来界定一个因子，用其尾部质量来界定另一个因子。利用引理 10.26

$$\sup_{t>0} g_u(t) \leqslant \frac{C}{\sqrt{u}} \qquad \text{for all large } u, \tag{10.46}$$

其中 $C > 0$为绝对常数。此外，

$$\int_{|z|>M} g_u(u + z\sqrt{2u})\, dz = \frac{1}{\sqrt{2u}} \mathbb{P}\Big(|X_u - u| > M\sqrt{2u}\Big) \leqslant \frac{2}{\sqrt{2u}} e^{-cM^2},$$

其中我们使用了卡方集中界，且当 u 足够大并且 M时，最后一个不等式对所有情形都成立。因此，对 $|x| \leqslant D_n$ 均匀地有，

$$\begin{aligned} &2\sqrt{2u} \int_{|z|>M} g_u(u + z\sqrt{2u})\, g_u(u + (z - \sqrt{2}x)\sqrt{2u})\, dz \\ &\leqslant 2\sqrt{2u} \cdot \sup_{t>0} g_u(t) \cdot \int_{|z|>M} g_u(u + z\sqrt{2u})\, dz \leqslant \frac{C}{\sqrt{u}}\, e^{-cM^2}. \end{aligned} \tag{10.47}$$

乘以 $\sqrt{u}$后，该尾部为 $\leqslant Ce^{-cM^2}$；通过取足够大的 M，可以使其任意小。

因此，只需在 $|z| \leqslant M$上分析（10.45）。

在 $|z| \leqslant M$ 和 $|x| \leqslant D_n$上，我们有 $|z - \sqrt{2}x| \leqslant M + \sqrt{2}D_n$。由于 $D_n = o(u^{1/6})$, 对于每个固定的 $M + \sqrt{2}D_n = o(u^{1/6})$ M也成立，因此引理 10.17 可均匀地应用：

$$g_u(u + z\sqrt{2u}) = \frac{1}{\sqrt{2u}} \phi(z)\big(1 + o(1)\big), \qquad g_u(u + (z - \sqrt{2}x)\sqrt{2u}) = \frac{1}{\sqrt{2u}} \phi(z - \sqrt{2}x)\big(1 + o(1)\big),$$

在 $|z| \leqslant M$ 和 $|x| \leqslant D_n$上一致地成立。代入（10.45）（限制于 $|z| \leqslant M$）可得

$$\begin{aligned} f_u(x\sqrt{u}) &= 2\sqrt{2u} \int_{|z|\leqslant M} \frac{1}{2u} \phi(z)\phi(z - \sqrt{2}x)\, dz \cdot (1 + o(1)) + O\left(\frac{1}{\sqrt{u}} e^{-cM^2}\right) \\ &= \frac{\sqrt{2}}{\sqrt{u}} \int_{|z|\leqslant M} \phi(z)\phi(z - \sqrt{2}x)\, dz \cdot (1 + o(1)) + O\left(\frac{1}{\sqrt{u}} e^{-cM^2}\right), \end{aligned} \tag{10.48}$$

对 $|x| \leqslant D_n$均匀成立。

先令 $u \to \infty$（固定 M），再令 $M \to \infty$。利用引理 10.25，$\int_{\mathbb{R}} \phi(z)\phi(z - \sqrt{2}x)\, dz = \frac{1}{\sqrt{2}} \phi(x)$ 对所有 $x \in \mathbb{R}$均成立，并且 $\int_{|z|>M} \phi(z)\phi(z - \sqrt{2}x)\, dz \leqslant \int_{|z|>M} \phi(z)\, dz \to 0$关于 x均匀地成立，由（10.48）得到

$$f_u(x\sqrt{u}) = \frac{1}{\sqrt{u}} \phi(x)\big(1 + o(1)\big),$$

对 $|x| \leqslant D_n$均匀成立，从而证明（10.24）。

对于第二个方程，我们希望证明其对所有 $b \in \mathbb{R}$均成立。对（10.44）关于 b 求导，得到

$$f_u'(b) = -4 \int_0^\infty g_u(t)\, g_u'(t - 2b)\, \mathbf{1}\{t - 2b > 0\}\, dt. \tag{10.49}$$

固定 $b = x\sqrt{u}$，令 $|x| \leqslant D_n$，并对 $t = u + z\sqrt{2u}$ 应用相同的变量替换，得到

$$f_u'(x\sqrt{u}) = -4\sqrt{2u} \int_{z>z_{u,x}} g_u(u + z\sqrt{2u})\, g_u'\big(u + (z - \sqrt{2}x)\sqrt{2u}\big)\, dz.$$

截断到 $|z| \leqslant M$，与之前完全相同：利用（10.46）和 $\sup_{t>0} |g_u'(t)| \leqslant C/u$，由引理 10.26以及卡方尾界可知，$|z| > M$ 的贡献为 $O\big(u^{-1}e^{-cM^2}\big)$，且对 $|x| \leqslant D_n$均匀成立。在$|z| \leqslant M$上，引理 10.17 给出

$$g_u(u + z\sqrt{2u}) = \frac{1}{\sqrt{2u}} \phi(z)\big(1 + o(1)\big), \qquad g_u'\big(u + (z - \sqrt{2}x)\sqrt{2u}\big) = \frac{1}{2u} \phi'(z - \sqrt{2}x)\big(1 + o(1)\big),$$

对 $|z| \leqslant M$ 和 $|x| \leqslant D_n$ 均一致成立。因此

$$f_u'(x\sqrt{u}) = -\frac{2}{u}\int_{|z|\leqslant M} \phi(z)\phi'(z-\sqrt{2}x)\,dz \cdot (1+o(1)) + O\left(\frac{1}{u}e^{-cM^2}\right),$$

关于 $|x| \leqslant D_n$ 一致地。令 $u \to \infty$，则 $M \to \infty$，并利用

$$\int_{\mathbb{R}} \phi(z)\phi'(z-\sqrt{2}x)\,dz = -\frac{1}{2}\phi'(x),$$

（这是通过对 $\int \phi(z)\phi(z-\sqrt{2}x)\,dz = \frac{1}{\sqrt{2}}\phi(x)$ 关于 x 求导得到的），我们得出

$$f_u'(x\sqrt{u}) = \frac{1}{u}\phi'(x)\big(1+o(1)\big),$$

对 $|x| \leqslant D_n$一致成立，从而证明了（10.25）。 □

引理 *10.19* 的证明。回顾一下，$B_u = \langle G_u,\ H_u\rangle = \sum_{i=1}^u G_i H_i$，其中 $G_u,\ H_u \sim N(0,\ I_u)$ 相互独立。对 G_u 进行条件化，我们有

$$B_u \mid G_u \sim N(0, \|G_u\|^2).$$

记 $V := \|G_u\|^2 \sim \chi^2_u$，可得

$$f_u(x) = \mathbb{E}[\varphi_V(x)] = \mathbb{E}\left[\frac{1}{\sqrt{2\pi V}}e^{-x^2/(2V)}\right], \tag{10.50}$$

其中 φ_v 表示 $N(0,\ v)$ 密度。

固定 $x \neq 0$。对于每个 $v > 0$,

$$\varphi_v(x) = \frac{1}{\sqrt{2\pi v}}e^{-x^2/(2v)} = \frac{1}{|x|\sqrt{\pi}}\left(\frac{x^2}{2v}\right)^{1/2} e^{-x^2/(2v)}.$$

由于函数 $t \mapsto t^{1/2}e^{-t}$ 在 $[0,\ \infty)$ 上有界，我们得到

$$\varphi_v(x) \leqslant \frac{C}{|x|} \qquad \text{for all } v > 0.$$

对（10.50）取期望，得到

$$f_u(x) \leqslant \frac{C}{|x|},$$

对 $u \geqslant 1$ 均匀成立。

接下来，对于 $x \neq 0$ 和 $v > 0$,

$$\varphi_v'(x) = -\frac{x}{v}\varphi_v(x),$$

so

$$|\varphi_v'(x)| = \frac{|x|}{\sqrt{2\pi}}v^{-3/2}e^{-x^2/(2v)} = \frac{2}{\sqrt{\pi}\,x^2}\left(\frac{x^2}{2v}\right)^{3/2} e^{-x^2/(2v)}.$$

由于 $t \mapsto t^{3/2}e^{-t}$ 在 $[0,\ \infty)$ 上也有界，因此

$$|\varphi_v'(x)| \leqslant \frac{C}{x^2} \qquad \text{for all } v > 0.$$

因此，对于每个固定的 $x \neq 0$，随机变量 $\varphi_V'(x)$ 被可积常数 C/x^2, 所控制，所以我们可以在期望下求导，因此 f_u 在 $\mathbb{R}\backslash\{0\}$ 上可微，且

$$f_u'(x) = \mathbb{E}[\varphi_V'(x)], \qquad \text{therefore} \qquad |f_u'(x)| \leqslant \frac{C}{x^2}.$$

□

10.4.1 引理的证明 10.25，10.26

引理的证明 *10.25*。 利用 $\phi(z) = (2\pi)^{-1/2} e^{-z^2/2}$ 我们写出

$$\int_{\mathbb{R}} \phi(z)\phi(z-\sqrt{2}x)\,dz = \frac{1}{2\pi}\int_{\mathbb{R}} \exp\Big(-\frac{z^2+(z-\sqrt{2}x)^2}{2}\Big)\,dz.$$

展开并配方：

$$-\frac{z^2+(z-\sqrt{2}x)^2}{2} = -(z^2-\sqrt{2}x\,z+x^2) = -\Big(z-\frac{x}{\sqrt{2}}\Big)^2 - \frac{x^2}{2}.$$

通过变量替换 $u = z - \frac{x}{\sqrt{2}}$，我们有 $-=$

$$\int_{\mathbb{R}} \exp\Big(-\Big(z-\frac{x}{\sqrt{2}}\Big)^2\Big)\,dz = \int_{\mathbb{R}} e^{-u^2}\,du = \sqrt{\pi}.$$

Thus

$$\int_{\mathbb{R}} \phi(z)\phi(z-\sqrt{2}x)\,dz = \frac{e^{-x^2/2}}{2\pi}\cdot\sqrt{\pi} = \frac{1}{2\sqrt{\pi}}e^{-x^2/2} = \frac{1}{\sqrt{2}}\cdot\frac{1}{\sqrt{2\pi}}e^{-x^2/2} = \frac{1}{\sqrt{2}}\phi(x),$$

如所述。 □

引理的证明 *10.26*。 首先，

$$\frac{d}{dt}\log g_u(t) = \frac{u/2-1}{t} - \frac{1}{2},$$

因此，当 $t < u-2$ 时，该导数为正；否则为负，这意味着：

$$\sup_{t>0} g_u(t) = g_u(u-2) = \frac{(u-2)^{u/2-1}e^{-(u-2)/2}}{2^{u/2}\Gamma(u/2)}.$$

利用斯特林近似的下界 $\Gamma(u/2) \geqslant \sqrt{2\pi}\,(u/2)^{u/2-1/2}e^{-u/2}$ 可得

$$\sup_{t>0} g_u(t) \leqslant \frac{e}{2\sqrt{\pi}}\cdot\frac{1}{\sqrt{u}}\Big(1-\frac{2}{u}\Big)^{u/2-1} \leqslant \frac{e}{2\sqrt{\pi}}\cdot\frac{1}{\sqrt{u}}.$$

接下来，

$$g_u'(t) = g_u(t)\Big(\frac{u/2-1}{t} - \frac{1}{2}\Big) = g_u(t)\cdot\frac{u-2-t}{2t},$$

因此

$$|g_u'(t)| = g_u(t)\cdot\frac{|u-2-t|}{2t}.$$

对于 $t \neq u-2$，最大化 $|g_u'(t)|$ 等价于最大化 $h(t) := t^{u/2-2}e^{-t/2}|u-2-t|,\quad t>0$. 在区间 $(0,u-2)$ 和 $(u-2,\infty)$上，函数 h 是 C^1，其临界点满足

$$(t-(u-2))^2 = 2(u-2).$$

此外，当 $t\to\infty$时， $h(t)\to 0$；对于 $u>2$，当 $t\downarrow 0$时也有 $h(t)\to 0$。由于 h 在 $(0,\infty)$上连续且非负，因此其全局最大值要么在边界处取得，要么在临界点处取得。在边界处的极限为 0，因此最大值在某个临界点处取得

$$t = u-2\pm\sqrt{2(u-2)}.$$

在任一此类点处， $|u-2-t| = \sqrt{2(u-2)}$；当 u 足够大时，我们有 $t \geqslant (u-2)/2$。因此

$$\sup_{t>0}|g_u'(t)| \leqslant \sup_{t>0} g_u(t)\cdot\frac{\sqrt{2(u-2)}}{2\cdot(u-2)/2} \leqslant \sup_{t>0} g_u(t)\cdot\frac{2}{\sqrt{u}}.$$

结合第一个界，$\sup_{t>0} g_u(t) \leqslant C/\sqrt{u}$ 可得 $\sup_{t>0}|g_u'(t)| \leqslant C/u$，其中已对普适常数 C进行调整。□

致谢

作者对 感谢 Hugo Koubbi 在本项目早期阶段进行的有益讨论 .

参考文献

[AGJ20] Gerard Ben Arous、Reza Gheissari 和 Aukosh Jagannath。《张量主成分分析的算法阈值》。《概率年刊》，48(4):2052–2087，2020。[AS65] Milton Abramowitz 和 Irene A Stegun。《数学函数手册：含公式、图表和数学表格》，第55卷。Courier Corporation，1965。[AW12] Rami Atar 和 Tsachy Weissman。《泊松信道中的互信息、相对熵与估计》。《IEEE信息论汇刊》，58(3):1302–1318，2012。[AWZ20] G´erard Ben Arous、Alexander S Wein 和 Ilias Zadik。《稀疏主成分分析中的自由能井与重叠间隙性质》。载于《学习理论会议》，第479–482页。PMLR，2020。[BEAH+22] Afonso S Bandeira、Ahmed El Alaoui、Samuel Hopkins、Tselil Schramm、Alexander S Wein 和 Ilias Zadik。《高维统计中的Franz-Parisi判据与计算权衡》。载于《神经信息处理系统进展》，35:33831–33844，2022。[BKM+19] Jean Barbier、Florent Krzakala、Nicolas Macris、L´eo Miolane 和 Lenka Zdeborov´a。《高维广义线性模型中的最优误差与相变》。《美国国家科学院院刊》，116(12):5451–5460，2019。[CGGV25] Alexandra Carpentier、Simone Maria Giancola、Christophe Giraud 和 Nicolas Verzelen。《通过近似正交归一基得到低阶下界》。arXiv预印本 arXiv:2509.09353，2025。[CMZZ25] Siyu Chen、Theodor Misiakiewicz、Ilias Zadik 和 Peiyuan Zhang。《优化的Franz-Parisi判据及其与SQ下界的等价性》。载于《神经信息处理系统第三十九届年会》，2025。[CSZ24] Zongchen Chen、Conor Sheehan 和 Ilias Zadik。《关于低温MCMC阈值：稀疏张量主成分分析、稀疏回归与几何规则》。arXiv预印本 arXiv:2408.00746，2024。[EGV25a] Bertrand Even、Christophe Giraud 和 Nicolas Verzelen。《潜变量模型中的计算下界：聚类、稀疏聚类与双聚类》。arXiv预印本 arXiv:2506.13647，2025。[EGV25b] Bertrand Even、Christophe Giraud 和 Nicolas Verzelen。《潜变量模型中的计算下界：聚类、稀疏聚类与双聚类》，2025。[FP95] Silvio Franz 和 Giorgio Parisi。《自旋玻璃亚稳态的构造方法》。《Journal de Physique I》，5(11):1401–1415，1995。[FP97] Silvio Franz 和 Giorgio Parisi。《耦合玻璃态系统的相图：一项平均场研究》。《物理评论快报》，79(13):2486，1997。[GSV05] Dongning Guo、Shlomo Shamai 和 Sergio Verd´u。《高斯信道中的互信息与最小均方误差》。《IEEE信息论汇刊》，51(4):1261–1282，2005。[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方面的威力》。载于2017年IEEE第58届计算机科学基础年会（FOCS），第720–731页，2017。[HS17] Samuel B Hopkins 和 David Steurer。《利用少量样本进行高效贝叶斯估计：社区检测及相关问题》。载于2017年IEEE第58届计算机科学基础年会（FOCS），第379–390页。IEEE，2017。

[KMW24] Dmitriy Kunisky、Cristopher Moore 和 Alexander S. Wein. 不变分布统计推断中的张量累积量，2024.[LG24] Yuetian Luo 和 Chao Gao. 通过低次多项式获得图子估计的计算下界。《统计年刊》，52(5)：2318–2348，2024.[LM17] Marc Lelarge 和 L´eo Miolane. 对称低秩矩阵估计的基本极限。载于学习理论会议，页码 1297–1301。PMLR，2017.[LWB21] Matthias L¨offler、Alexander S. Wein 和 Afonso S. Bandeira. 计算高效的稀疏聚类，2021.[LZ23] Yuetian Luo 和 Anru R. Zhang. 具有植入结构的张量聚类：统计最优性与计算极限，2023.[McC87] Peter McCullagh. 统计学中的张量方法。1987.[MNWS+23] Elchanan Mossel、Jonathan Niles-Weed、Youngtak Sohn、Nike Sun 和 Ilias Zadik. 植入子图推断中的尖锐阈值。载于第三十六届学习理论年度会议，页码 5573–5577。PMLR，2023.[MOS13] Wilhelm Magnus、Fritz Oberhettinger 和 Raj Pal Soni. 数理物理特殊函数的公式与定理，第52卷。Springer Science & Business Media，2013.[MR14] Andrea Montanari 和 Emile Richard. 张量主成分分析的统计模型，2014.[MW24] Andrea Montanari 和 Alexander S. Wein. 秩一矩阵估计中近似消息传递与低次多项式的等价性，2024.[NWZ20] Jonathan Niles-Weed 和 Ilias Zadik. 稀疏张量主成分分析中的全有或全无现象。《神经信息处理系统进展》，33：17674–17684，2020.[RP16] Galen Reeves 和 Henry D Pfister. 高斯矩阵压缩感知的副本对称预测是精确的。载于2016年IEEE国际信息论研讨会（ISIT），页码 665–669。IEEE，2016.[RXZ19] Galen Reeves、Jiaming Xu 和 Ilias Zadik. 稀疏线性回归中的全有或全无现象。载于学习理论会议，页码 2652–2663。PMLR，2019.[SW22] Tselil Schramm 和 Alexander S. Wein. 由低次多项式进行估计的计算障碍。《统计年刊》，50(3)，2022年6月.[SW25] Youngtak Sohn 和 Alexander S. Wein. 使用低次多项式进行估计时的尖锐相变，2025.[VdV00] Aad W Van der Vaart. 渐近统计，第3卷。剑桥大学出版社，2000.[Ver18] Roman Vershynin. 高维概率：数据科学中的引言与应用，第47卷。剑桥大学出版社，2018.[VGNA20] Mariia Vladimirova、St´ephane Girard、Hien Nguyen 和 Julyan Arbel. 次韦布尔分布：将次高斯和次指数性质推广到更重尾的分布。《Stat》，9(1)，2020年1月.[YC17] Zhen-Hang Yang 和 Yu-Ming Chu. 关于第二类修正贝塞尔函数的逼近。《不等式与应用杂志》，2017(1)：41，2017年2月.[ZK16] Lenka Zdeborov´a 和 Florent Krzakala. 推断的统计物理：阈值与算法。《物理学进展》，65(5)：453–552，2016。

A 未能 预测低阶困难性的淬火 FP 势

自然会想知道，原始淬火 FP 势的单调性 $\mathcal{F}_\lambda$ （1）是否等价于一族 GAM 的低次 MMSE 下界。换句话说：

能否直接根据淬火 *FP* 势的单调性行为推导出低阶 *MMSE* 的结果？

虽然我们的结果确立了这一联系适用于退火 FP 势，但传统的物理直觉认为，同样的结论也应适用于淬火势，因为退火势通常仅被定义为淬火能景的一个易于处理的替代指标。本研究中的许多结果仍有可能推广到淬火情形，例如直接确立退火势与淬火势的单调性一致。然而，本附录旨在提出一个发人深省的反例：我们证明，尽管退火 FP 势的单调性与低次 MMSE 下界相符，淬火 FP 势的单调性却并非如此。两种势函数之间的这一差异值得关注；至于为何退火势在预测算法困难性方面能够“胜过”淬火势，我们将这一问题留作未来研究的一个重要课题。

我们明确构造并分析了这个反例。在这一设定中，当低阶 MMSE 平凡时，退火 FP 恰好在“0 附近”递增（即退火物理困难）；而当低阶 MMSE 优于平凡 MSE 时，退火 FP 恰好开始在“0 附近”递减（即退火物理容易）。相比之下，即使在低阶 MMSE 已经严格优于平凡 MSE 的区域内，淬火 FP 在“0 附近”仍保持非递减（即并非淬火物理容易）。

模型反例 该反例适用于以下“截断”拉*ademacher” 3*-张量稀疏主成分分析模型。对于 $k = n^{\beta+o(1)}$ 其中 $\beta \in (0,\ 1/2)$, 我们选择 $X = \mathrm{vec}(v^{\otimes 3})$, 对于 $v \in \mathbf{R}^n$ ，其生成方式如下。令 $u \in \mathbf{R}^n$ 的元素独立同分布为 $\mathrm{Rad}(k/n),$，这意味着对于所有 $i = 1,\ \ldots,\ N$， $u_i = 1$ 以概率 $k/(2n)$, $u_i = -1$ 以概率 $k/(2n)$ ，否则为 $u_i = 0$。如果 $\|u\|_0 \in [k/2, 2k],$，我们设 $v = u.$。否则我们设 $v = 1_{[k]}$ （前 k 个元素的示性函数）。

注 A.1 （解释“截断”并概述全文结构）。 该模型与第 3.1.2 节讨论的 Rademacher 稀疏张量 PCA 模型几乎完全相同，唯一的区别在于，我们将信号截断为v 至 $1_{[k]}$ ，只要 $\|v\|_0 \notin [k/2, 2k]$。Bernstein 不等式的标准应用表明，事件$\|v\|_0 \in [k/2, 2k]$ 发生的概率为 $1 - e^{-\Theta(k)}$。因此，截断模型和原始模型具有相同的“算法”阈值（见第 A.1 节）、相同的“低次”阈值（见第 A.2 节），以及相同的退火 Franz-Parisi 势行为（见第 A.2 节）。特别是，结合我们在第 3.1.2 节中的等价定理，对于截断模型，退火 FP 势的单调性仍然等价于低阶 MMSE 界。

引入截断纯粹是出于技术上的便利，因为这简化了对淬火 Franz-Parisi 势的分析。为完成反例，我们证明淬火 FP 势的单调性与低次 MMSE 下界并不一致（见第 A 节）。具体而言，在低次“易”区间的很大一部分范围内，淬火 FP 势仍然是非单调的。

A.1 算法阈值

~在本节中，我们证明对于 $\lambda = \Theta(1)$ ，可以通过一种（简单的）对角阈值法多项式时间方法实现高概率精确恢复（因此其 MMSE 的阶小于平凡均方误差；见算法 1）。特别地，先验中的“截断”不会影响对稀疏 Rademacher 三阶张量 PCA 模型使用对角阈值法的成功率。

本小节的主要结果如下。

引理 A.2（对角阈值法可恢复 v （包括符号）。设 Y 由截断稀疏三阶张量 *PCA* 模型生成。对于每个 $i \in [n]$, 令 $d_i := Y_{iii}$。定义对角阈值估计量 $\widehat{v} \in \{-1,\ 0,\ 1\}^n$ 如下：

$$\widehat{v}_i := \mathrm{sign}(d_i)1\{|d_i| \geqslant \tau\}, \qquad i = 1, \ldots, n. \tag{A.1}$$

算法 1 用于支撑恢复的对角阈值法

要求：张量 $Y \in \mathbb{R}^{n^3}$，阈值 $\tau > 0$1：计算 $i = 1, \ldots, n$ 对应的对角项 $d_i \leftarrow Y_{iii}$2：输出 $\widehat{v} \leftarrow$ 的符号 $(d_i)1\{|d_i| \geqslant \tau\}$

令 $\tau := \sqrt{6}$对数 n并假设 $\lambda \geqslant 2\tau$。那么

$$\mathbb{P}(\widehat{v} \neq v) \leqslant 2n^{-2} + 4k\,n^{-3} + 4k\,n^{-27}. \tag{A.2}$$

特别地，若对于任意 $\beta \in (0, \quad k = n^{\beta + o(1)}\ 1/2),$，则 $\mathbb{P}(\widehat{v} \neq v) = o(1)$。

此引理的证明推迟到第 B 节。

A.2 退火 FP 势与低阶 MMSE 的等价性

我们的下一个论断是，退火 FP 势的单调性等价于截断的稀疏三阶张量 PCA 模型的低阶 MMSE 下界。提醒读者，对于第3.1.2 节中的原始（“未截断”）稀疏三阶张量 PCA 模型，这种等价性已在定理 3.2 中讨论；具体而言，由于原始模型满足假设 2.10，该结论可由定理 2.13 和2.11 推出。

由于将截断引入先验分布会带来技术上的复杂性，我们证明截断稀疏三阶张量 PCA 模型的等价性时，并不直接验证假设 2.10，而是证明 (a) 退火 FP 势的单调性和 (b) 低阶 MMSE 在原始稀疏三阶张量 PCA 模型与截断稀疏三阶张量 PCA 模型之间仅相差 $1 + o(1)$ 因子。由于原始模型满足等价性，因此该等价性也适用于截断模型。

A.2.1 原始模型与截断模型之间低阶 MMSE 的等价性

首先，我们说明截断的 3-稀疏张量 PCA 模型与原始模型的低阶 MMSE 几乎等价。设 P 为 3-稀疏张量情形下信号 $(u^{\otimes 3}) \in \mathbb{R}^N$ 的未截断先验分布，$N = n^3$，并令 $\widetilde{P}$ 表示截断先验分布。

命题 A.3. 存在一个绝对常数 $C > 0$，使得对于任意满足 $D = O((\text{对数}\, n)^h)$的 D，其中常数 $0 \leqslant h < 2$ 和 $\lambda = O(1)$,

$$\left|\mathrm{MMSE}_{\widetilde{P}}^{\leqslant D}(\lambda) - \mathrm{MMSE}_{P}^{\leqslant D}(\lambda)\right| \leqslant e^{-\Theta(k)}.$$

证明。我们按以下步骤完成证明。

步骤1：耦合并化归到坏事件。我们采用两个先验分布之间显然的耦合。首先，抽取 $\widetilde{X} = \mathsf{vec}(u^{\otimes 3}) \sim \widetilde{P}$。记 $E_1(\widetilde{X}) = \{\|u\|_0 \in [k/2, 2k]\}$ 为依赖于 $\widetilde{X}$ 的可测事件。固定确定性向量 $x_0 = \mathsf{vec}(1_{[k]}^{\otimes 3}) \in \mathbb{R}^N$。那么，对于截断信号

$$X := \widetilde{X}\mathbf{1}_{E_1(\widetilde{X})} + x_0\mathbf{1}_{E_1(\widetilde{X})^c},$$

我们有 $X \sim P$。令 $Z \sim \mathcal{N}(0, I_N)$ 相互独立，并定义

$$\widetilde{Y} := \sqrt{\lambda}\widetilde{X} + Z, \qquad Y := \sqrt{\lambda}X + Z.$$

根据构造，$(X, Y) = (\widetilde{X}, \widetilde{Y})$ 在 $E_1(\widetilde{X})$ 上成立。因此，对于任意 $f \in (\mathbb{R}[\cdot]_{\leqslant D})^N$,

$$\begin{aligned}
\left|\mathbb{E}\|f(Y) - X\|^2 - \mathbb{E}\|f(\widetilde{Y}) - \widetilde{X}\|^2\right| &= \left|\mathbb{E}\Big[\big(\|f(Y) - X\|^2 - \|f(\widetilde{Y}) - \widetilde{X}\|^2\big)\mathbf{1}_{E_1(\widetilde{X})^c}\Big]\right| \\
&\leqslant \mathbb{E}\Big[\big(\|f(Y) - X\|^2 + \|f(\widetilde{Y}) - \widetilde{X}\|^2\big)\mathbf{1}_{E_1(\widetilde{X})^c}\Big] \\
&\leqslant 2\mathbb{E}\Big[\big(\|f(Y)\|^2 + \|f(\widetilde{Y})\|^2 + \|X\|^2 + \|\widetilde{X}\|^2\big)\mathbf{1}_{E_1(\widetilde{X})^c}\Big].
\end{aligned} \tag{A.3}$$

现在我们继续对这四项进行界定：

$$\mathbb{E}[\|f(Y)\|^2\mathbf{1}_{E_1^c}], \quad \mathbb{E}[\|f(\widetilde{Y})\|^2\mathbf{1}_{E_1^c}], \quad \mathbb{E}[\|X\|^2\mathbf{1}_{E_1^c}], \quad \mathbb{E}[\|\widetilde{X}\|^2\mathbf{1}_{E_1^c}].$$

步骤 2：限制估计量类 f。首先，我们记录以下基本事实。对于任意随机向量 U，V和估计量 $f(V)$，若 $\mathbb{E}\|f(V)-U\|^2 \leqslant \mathbb{E}\|U\|^2$，则根据三角不等式，

$$\|f(V)\| \leqslant \|f(V)-U\| + \|U\|,$$

因此

$$\|f(V)\|^2 \leqslant 2\|f(V)-U\|^2 + 2\|U\|^2.$$

取期望可得

$$\mathbb{E}\|f(V)\|^2 \leqslant 4\mathbb{E}\|U\|^2.$$

在界定低阶 MMSE 时，由于最优估计量总是优于零估计量，我们可以将估计量限制为满足

$$\mathbb{E}\|f(Y)-X\|^2 \leqslant \mathbb{E}\|X\|^2.$$

因此，

$$\mathbb{E}\|f(Y)\|^2 \leqslant 4\mathbb{E}\|X\|^2 \leqslant Ck^3.$$

类似地，在 $\widetilde{P}$ 下，我们也有 $\mathbb{E}\|f(\widetilde{Y})\|^2 \leqslant Ck^3$.

~步骤 3：坏事件的概率。令 $\delta := \mathrm{P}_{\widetilde{P}}(E_1(X)^c)$。由于 $\|u\|_0 \sim$二项分布$(n,\ k/n)$是均值为 k、方差至多为 k 的次高斯变量，Bernstein 不等式表明存在一个绝对常数 $C>0$，使得 $\delta \leqslant e^{-2ck}$。

步骤 4：信号项的界。在 E_1^c 上，我们有 $X = x_0$，因此

$$\mathbb{E}[\|X\|^2 \mathbf{1}_{E_1^c}] = \delta k^3. \tag{A.4}$$

此外，由于 $\|\widetilde{X}\|^2 = \|u\|_0^3$ 且 $\mathbb{E}[\|\widetilde{X}\|^4] \leqslant C^2k^6$ (参见例如 [Ver18，命题 2.5。2])，由柯西–施瓦茨不等式，

$$\mathbb{E}[\|\widetilde{X}\|^2 \mathbf{1}_{E_1^c}] \leqslant \sqrt{\mathbb{E}[\|\widetilde{X}\|^4]\mathbb{P}(E_1^c)} \leqslant Ck^3 e^{-ck}. \tag{A.5}$$

步骤 5：$\mathbb{E}[\|f(Y)\|^2 \mathbf{1}_{E_1^c}]$的界。在 E_1^c 上，$Y = \sqrt{\lambda}x_0 + Z$，因此

$$\mathbb{E}[\|f(Y)\|^2 \mathbf{1}_{E_1^c}] = \delta\, \mathbb{E}\|f(Y)\|^2 \leqslant Ck^3 e^{-2ck}. \tag{A.6}$$

步骤 6：$\mathbb{E}[\|f(\widetilde{Y})\|^2 \mathbf{1}_{E_1^c}]$的界。剩下要控制的唯一非平凡项是 $\mathbb{E}[\|f(\widetilde{Y})\|^2 \mathbf{1}_{E_1^c}]$。

步骤 6a：通过埃尔米特展开得到上界。对 u 进行条件化后，我们首先证明

$$\mathbb{E}[\|f(\widetilde{Y})\|^2 \mathbf{1}_{E_1^c}] = \mathbb{E}_u\left[\mathbf{1}_{E_1(\widetilde{X})^c}\mathbb{E}_Z[\|f(Z+\sqrt{\lambda}\widetilde{X})\|^2]\right] \leqslant \mathbb{E}_Z[\|f(Z)\|^2]\sum_{j=0}^{D}\binom{D}{j}\frac{1}{j!}\lambda^j \mathbb{E}[\|u\|_0^{3j}\mathbf{1}_{E_1^c}]. \tag{A.7}$$

我们通过将 f 按多元埃尔米特基展开为 $f(z) = \sum_{|\alpha|\leqslant D}\widehat{f}_\alpha H_\alpha(z)$ 来证明这个恒等式，由此可得 $\mathbb{E}_Z[\|f(Z)\|^2] = \sum_{|\alpha|\leqslant D}\widehat{f}_\alpha^2 \alpha!$ 。

令 $\mu = \sqrt{\lambda}\widetilde{X}$。应用命题7.3，可得

$$f(Z+\mu) = \sum_{|\alpha|\leqslant D}\widehat{f}_\alpha \sum_{\gamma\leqslant\alpha}\binom{\alpha}{\gamma}\mu^{\alpha-\gamma}H_\gamma(Z);$$

并根据引理7.5，

$$\mathbb{E}_Z[\|f(Z+\mu)\|^2] = \sum_{|\gamma|\leqslant D}\gamma!\left\|\sum_\beta \widehat{f}_{\gamma+\beta}\binom{\gamma+\beta}{\beta}\mu^\beta\right\|^2.$$

我们对关于 β 的内层求和应用柯西–施瓦茨不等式：

$$\left\|\sum_{\beta}\widehat{f}_{\gamma+\beta}\sqrt{(\gamma+\beta)!}\frac{\binom{\gamma+\beta}{\beta}}{\sqrt{(\gamma+\beta)!}}\mu^{\beta}\right\|^{2}\leqslant\left(\sum_{\beta}\|\widehat{f}_{\gamma+\beta}\|^{2}(\gamma+\beta)!\right)\left(\sum_{\beta}\frac{\binom{\gamma+\beta}{\beta}^{2}}{(\gamma+\beta)!}\mu^{2\beta}\right).$$

两边同时乘以 $\gamma!$，并利用阶乘恒等式 $\gamma!\frac{\binom{\gamma+\beta}{\beta}^{2}}{(\gamma+\beta)!}=\binom{\gamma+\beta}{\beta}\frac{1}{\beta!}$，得到

$$\mathbb{E}_{Z}[\|f(Z+\mu)\|^{2}]\leqslant\sum_{|\gamma|\leqslant D}\left(\sum_{\beta}\|\widehat{f}_{\gamma+\beta}\|^{2}(\gamma+\beta)!\right)\left(\sum_{\beta}\binom{\gamma+\beta}{\beta}\frac{\mu^{2\beta}}{\beta!}\right),$$

将变量换回 $\alpha=\gamma+\beta$ 后，对 γ 和 β 求和，其中 $|\gamma+\beta|\leqslant D$ 等价于对 $|\alpha|\leqslant D$ 和 $\beta\leqslant\alpha$ 求和：

$$\mathbb{E}_{Z}[\|f(Z+\mu)\|^{2}]\leqslant\sum_{|\alpha|\leqslant D}\|\widehat{f}_{\alpha}\|^{2}\alpha!\sum_{\beta\leqslant\alpha}\binom{\alpha}{\beta}\frac{\mu^{2\beta}}{\beta!}.$$

从组合意义上看，$\binom{\alpha}{\beta}=\prod_{i=1}^{N}\binom{\alpha_i}{\beta_i}\leqslant\binom{|\alpha|}{|\beta|}\leqslant\binom{D}{|\beta|}$。令 $j=|\beta|$。按次数 j 对内层求和分组，得到

$$\sum_{\beta\leqslant\alpha}\binom{\alpha}{\beta}\frac{\mu^{2\beta}}{\beta!}\leqslant\sum_{j=0}^{|\alpha|}\binom{D}{j}\sum_{|\beta|=j}\frac{\mu^{2\beta}}{\beta!}$$

根据多项式定理，$\sum_{|\beta|=j}\frac{j!}{\beta!}\mu^{2\beta}=(\sum_{i=1}^{N}\mu_i^{2})^{j}=\|\mu\|^{2j}$。因此，对于任意 α，内层和均有统一上界 $\sum_{j=0}^{D}\binom{D}{j}\frac{1}{j!}\|\mu\|^{2j}$。将其提取出来，即可得到基准范数：

$$\mathbb{E}_{Z}[\|f(Z+\mu)\|^{2}]\leqslant\left(\sum_{|\alpha|\leqslant D}\|\widehat{f}_{\alpha}\|^{2}\alpha!\right)\sum_{j=0}^{D}\binom{D}{j}\frac{1}{j!}\|\mu\|^{2j}=\mathbb{E}_{Z}[\|f(Z)\|^{2}]\sum_{j=0}^{D}C_{j}(D)\|\mu\|^{2j}\tag{A.8}$$

其中 $C_j(D)=\binom{D}{j}\frac{1}{j!}$，这证明了（A.7）。

回顾 $\|\mu\|^{2}=\lambda\|\widetilde{X}\|^{2}=\lambda\|u\|_{0}^{3}$，其中 $\|u\|_0\sim$ 二项分布$(n,\ k/n)$。类似地，根据 [Ver18，命题 2.5。2]，对于某个绝对常数 $C>0$，对每个 $j\leqslant D$，$\mathbb{E}[\|u\|_{0}^{6j}]\leqslant(Ck)^{6j}$。然后利用柯西–施瓦茨不等式，

$$\mathbb{E}[\|u\|_{0}^{3j}\mathbf{1}_{E_1(\widetilde{X})^{c}}]\leqslant\sqrt{\mathbb{E}[\|u\|_{0}^{6j}]\mathbb{P}(E_1(\widetilde{X})^{c})}=\sqrt{\mathbb{E}[\|u\|_{0}^{6j}]}\sqrt{\delta}\leqslant(Ck)^{3j}e^{-ck},$$

（A.7）可得

$$\mathbb{E}[\|f(\widetilde{Y})\|^{2}\mathbf{1}_{E_1^{c}}]\leqslant\mathbb{E}_{Z}[\|f(Z)\|^{2}]\cdot\frac{1}{(D-1)!}(C\lambda k)^{3D}e^{-ck}.$$

步骤 6b：关于 $\mathbb{E}\|f(\widetilde{Y})\|^{2}$ 和 $\mathbb{E}_{Z}[\|f(Z)\|^{2}]$ 的下界。最后，我们需要用 $\mathbb{E}\|f(\widetilde{Y})\|^{2}$ 对 $\mathbb{E}_{Z}[\|f(Z)\|^{2}]$ 给出上界，并使用步骤1。

对于每个多重指标 γ，令 $A_{\gamma}(\mu)=\sum_{\beta>0}\widehat{f}_{\gamma+\beta}\binom{\gamma+\beta}{\beta}\mu^{\beta}$ 表示严格高阶的移位分量。然后根据詹森不等式，

$$\mathbb{E}_{\mu}\left[\|\widehat{f}_{\gamma}+A_{\gamma}(\mu)\|^{2}\right]\geqslant\|\widehat{f}_{\gamma}+\mathbb{E}[A_{\gamma}(\mu)]\|^{2}.\tag{A.9}$$

利用柯西–施瓦茨不等式，对于任意 $\eta>0$、$(a+b)^{2}\geqslant(1+\eta)^{-1}a^{2}-\eta^{-1}b^{2}$。因此，我们得到

$$\mathbb{E}[\|f(Z+\mu)\|^{2}]\geqslant(1+\eta)^{-1}\sum_{|\gamma|\leqslant D}\gamma!\|\widehat{f}_{\gamma}\|^{2}-\eta^{-1}\sum_{|\gamma|\leqslant D}\gamma!\|\mathbb{E}[A_{\gamma}(\mu)]\|^{2}.$$

现在定义矩 $m_\beta = \mathbb{E}[\mu^\beta]$。在 β 求和中使用柯西–施瓦茨不等式，可得

$$\|\mathbb{E}[A_\gamma(\mu)]\|^2 = \left\| \sum_{\beta>0} \widehat{f}_{\gamma+\beta} \binom{\gamma+\beta}{\beta} m_\beta \right\|^2 \leqslant \left(\sum_{\beta>0} (\gamma+\beta)! \|\widehat{f}_{\gamma+\beta}\|^2 \right) \left(\sum_{\beta>0} \frac{\binom{\gamma+\beta}{\beta}^2 m_\beta^2}{(\gamma+\beta)!} \right).$$

乘以 $\gamma!$，并利用

$$\gamma! \frac{\binom{\gamma+\beta}{\beta}^2}{(\gamma+\beta)!} = \binom{\gamma+\beta}{\beta} \frac{1}{\beta!},$$

我们得到

$$\begin{aligned} \sum_{|\gamma|\leqslant D} \gamma! \|\mathbb{E}[A_\gamma(\mu)]\|^2 &\leqslant \left(\sum_{\beta>0} (\gamma+\beta)! \|\widehat{f}_{\gamma+\beta}\|^2 \right) \left(\sum_{\beta>0} \binom{\gamma+\beta}{\beta} \frac{m_\beta^2}{\beta!} \right) \leqslant \left(\sup_{|\alpha|\leqslant D} \sum_{0<\beta\leqslant\alpha} \binom{\alpha}{\beta} \frac{m_\beta^2}{\beta!} \right) \sum_{|\alpha|\leqslant D} \alpha! \|\widehat{f}_\alpha\|^2 \\ &:= S_D \sum_{|\alpha|\leqslant D} \alpha! \|\widehat{f}_\alpha\|^2 = S_D \cdot \mathbb{E}_Z[\|f(Z)\|^2], \end{aligned}$$

其中记

$$S_D = \sup_{|\alpha|\leqslant D} \sum_{0<\beta\leqslant\alpha} \binom{\alpha}{\beta} \frac{m_\beta^2}{\beta!}.$$

结合前一个不等式，我们得到

$$\mathbb{E}[\|f(Z+\mu)\|^2] \geqslant ((1+\eta)^{-1} - \eta^{-1} S_D) \mathbb{E}_Z[\|f(Z)\|^2]. \tag{A.10}$$

令 $p = k/n$。μ 的坐标为 $\mu_{abc} = \sqrt{\lambda} u_a u_b u_c$，对每个 $a,\ b,\ c \in [n]$。令 $e_i(\beta)$ 为 μ^β 中 u_i 的指数，并定义

$$r(\beta) = |\{i : e_i(\beta) > 0\}|.$$

由于

$$\mathbb{E}[u_i^r] = \begin{cases} 1, & r = 0, \\ 0, & r \text{ odd}, \\ p, & r \geqslant 2 \text{ even}, \end{cases}$$

我们得到

$$m_\beta = \lambda^{|\beta|/2} p^{r(\beta)} \mathbf{1}_{\{e_i(\beta) \text{ all even}\}}.$$

我们根据 $r(\beta) = 1$ 或 $r(\beta) \geqslant 2$ 的情况，将定义 S_D 的和拆分为：

$$S_D \leqslant S_D^{(1)} + S_D^{(\geqslant 2)},$$

其中

$$S_D^{(1)} := \sup_{|\alpha|\leqslant D} \sum_{\substack{0<\beta\leqslant\alpha \\ r(\beta)=1}} \binom{\alpha}{\beta} \frac{m_\beta^2}{\beta!}, \qquad S_D^{(\geqslant 2)} := \sup_{|\alpha|\leqslant D} \sum_{\substack{0<\beta\leqslant\alpha \\ r(\beta)\geqslant 2}} \binom{\alpha}{\beta} \frac{m_\beta^2}{\beta!}.$$

假设 $r(\beta) = 1$ 和 $m_\beta \neq 0$。那么存在唯一的 $i \in [n]$ 使得 $e_i(\beta) > 0$ 且对所有 $j \neq i$都有 $e_j(\beta) = 0$。这意味着，β支撑集中的每个张量坐标 (a,b,c) 都必须满足 $a = b = c = i$。因此，β必须支撑在单个纯对角坐标 (i,i,i) 上，也就是说，对某个整数 $b \geqslant 1$，有 $\beta = be_{(i,i,i)}$，其中 $e_{(i,i,i)}$ 表示相应的坐标基向量。此外，要使 m_β 非零，所有指数 $e_j(\beta)$ 都必须是偶数；此处 $e_i(\beta) = 3b$，因此 b 必为偶数。因此，满足 $r(\beta) = 1$ 的非零项恰好对应于多重指标 $\beta = 2qe_{(i,i,i)}$, $q \geqslant 1$。对于这样的 β，有 $m_\beta^2 = \lambda^{2q} p^2$。固定满足 $|\alpha| \leqslant D$的 α。那么

$$\sum_{\substack{0<\beta\leqslant\alpha \\ r(\beta)=1}} \binom{\alpha}{\beta} \frac{m_\beta^2}{\beta!} = p^2 \sum_{i=1}^n \sum_{\substack{2q\leqslant\alpha_{(i,i,i)} \\ q\geqslant 1}} \binom{\alpha_{(i,i,i)}}{2q} \frac{\lambda^{2q}}{(2q)!}.$$

对每个固定的 $q \geqslant 1$，利用

$$\sum_{i=1}^{n} \binom{\alpha_{(i,i,i)}}{2q} \leqslant \binom{\sum_{i=1}^{n} \alpha_{(i,i,i)}}{2q} \leqslant \binom{|\alpha|}{2q} \leqslant \binom{D}{2q},$$

我们得到

$$\sum_{\substack{0<\beta\leqslant\alpha \\ r(\beta)=1}} \binom{\alpha}{\beta} \frac{m_\beta^2}{\beta!} \leqslant p^2 \sum_{q=1}^{\lfloor D/2 \rfloor} \binom{D}{2q} \frac{\lambda^{2q}}{(2q)!}.$$

对 $|\alpha| \leqslant D$ 取上确界，得到

$$S_D^{(1)} \leqslant p^2 \sum_{q=1}^{\lfloor D/2 \rfloor} \binom{D}{2q} \frac{\lambda^{2q}}{(2q)!} \leqslant p^2 \sum_{q=1}^{\lfloor D/2 \rfloor} \frac{(D\lambda)^{2q}}{((2q)!)^2} \leqslant p^2 I_0(2\sqrt{D\lambda}),$$

其中， $I_0(z)$ 是附录 9 中定义的第一类修正贝塞尔函数。利用修正贝塞尔函数的积分表示， $I_0(z) = \frac{1}{\pi}\int_0^\pi \exp(z\cos\theta)d\theta$。由于 $\cos\ \theta \leqslant 1$ 在整个积分区间上，积分被 $\exp(z)$ 所界。因此， $I_0(z) \leqslant \exp(z)$，且 $S_D^{(1)} \leqslant p^2\exp(2\sqrt{D\lambda})$。

由于 $p = k/n,\ k = O(\sqrt{n}),\ D = O((\log n)^h)$ ，其中 $h < 2$、 $\lambda = O(1)$，我们可知 $\exp(2\sqrt{D\lambda}) = n^{o(1)}$，因此$S_D^{(1)} = o(1)$。

如果 $r(\beta) \geqslant 2$ 且 $m_\beta \neq 0$，那么 $m_\beta^2 = \lambda^{|\beta|}p^{2r(\beta)} \leqslant \lambda^{|\beta|}p^4$。因此，对每个满足 $|\alpha| \leqslant D$ 的 α ，

$$S_D^{(2)} = \sum_{\substack{0<\beta\leqslant\alpha \\ r(\beta)\geqslant 2}} \binom{\alpha}{\beta} \frac{m_\beta^2}{\beta!} \leqslant p^4 \sum_{0<\beta\leqslant\alpha} \binom{\alpha}{\beta} \frac{\lambda^{|\beta|}}{\beta!} = \widetilde{O}(p^2) = o(1).$$

上述论证表明 $S_D = o(1)$，因此当 n足够大时， $S_D \leqslant 0.4$。通过在 (A.10) 中选择 $\eta = 0.99$ ，

$$\mathbb{E}_P\|f(\widetilde{Y})\|^2 \geqslant 0.1\mathbb{E}_Z[\|f(Z)\|^2].$$

代回 (A.7)，并结合步骤1，我们得到如下结论：

$$\mathbb{E}[\|f(\widetilde{Y})\|^2 \mathbf{1}_{E_1^c}] \leqslant \frac{10}{(D-1)!} k^3 (C\lambda k)^{3D} e^{-ck} = e^{-\Theta(k)}. \tag{A.11}$$

步骤7：结论。根据不等式 (A.3)、(A.4)、(A.5)、(A.6) 和 (A.11)，我们有

$$\begin{aligned}
\left|\mathrm{MMSE}_P^{\leqslant D}(\lambda) - \mathrm{MMSE}_{\widetilde{P}}^{\leqslant D}(\lambda)\right| &= \left| \inf_{f\in\mathbb{R}^N[Y]_{\leqslant D}} \mathbb{E}\left[\|f(Y) - X\|^2\right] - \inf_{f\in\mathbb{R}^N[\widetilde{Y}]_{\leqslant D}} \mathbb{E}\left[\|f(\widetilde{Y}) - \widetilde{X}\|^2\right] \right| \\
&\leqslant \sup_{f\in\mathbb{R}^N_{\leqslant D}, \mathbb{E}\|f(\widetilde{Y})\|^2\leqslant Ck^3, \mathbb{E}\|f(Y)\|^2\leqslant Ck^3} \left| \mathbb{E}\|f(Y) - X\|^2 - \mathbb{E}\|f(\widetilde{Y}) - \widetilde{X}\|^2 \right| \\
&\leqslant 2 \sup_{f\in\mathbb{R}^N_{\leqslant D}, \mathbb{E}\|f(\widetilde{Y})\|^2\leqslant Ck^3, \mathbb{E}\|f(Y)\|^2\leqslant Ck^3} \left| \mathbb{E}\left[\left(\|f(Y)\|^2 + \|f(\widetilde{Y})\|^2 + \|X\|^2 + \|\widetilde{X}\|^2\right) \mathbf{1}_{E_1(\widetilde{X})^c} \right] \right| \\
&\leqslant Ck^3 (C\lambda k)^{3D} e^{-ck} = e^{-\Theta(k)}.
\end{aligned}$$

□

A.2.2 原始模型与截断模型之间退火 FP 势单调性的等价性

接下来，我们证明对先验分布进行截断不会改变退火 FP 势的单调性。具体而言，我们通过证明以下结论来实现这一点：从截断先验分布中独立同分布抽取的两个样本 $\langle X, X'\rangle$ 的重叠随机变量 X, X' 的对数概率质量函数，与原始先验分布的对数概率质量函数仅相差一个乘法因子 $o(1)$ 。

我们首先介绍一些符号说明。对于 $v, v' \in \{-1,0,1\}^n$，令 $S := \langle v, v' \rangle$, 并定义

$$E_2 = E_2(v, v') := \Big\{ \|v\|_0 \in [k/2, 2k] \Big\} \cap \Big\{ \|v'\|_0 \in [k/2, 2k] \Big\}.$$

~注意，在此符号约定下， v, v' 上的原始先验分布为 P，其中 v, v' 是独立同分布的，且具有独立同分布的 Rad 分布(k/n)坐标，与截断后的 P 的关系为:

$$\mathbb{P}_P(\,\cdot\,) := \mathbb{P}(\,\cdot \mid E_2).$$

我们将截断模型和原始模型的退火 FP 势分别记为

$$\mathcal{F}^{\tilde{P}}_{\mathrm{ann},\lambda}(q), q \in \mathbb{R}, \text{ and, } \mathcal{F}^{P}_{\mathrm{ann},\lambda}(q), q \in \mathbb{R}.$$

我们证明关于它们的以下结论。

定理 A.4. 固定 $\beta \in (0, 1/2)$, 令 $k = n^{\beta + o(1)}$。对于任意 $n^{-C} \leqslant \lambda \leqslant n^C$，存在常数 $C > 0$，使得对任意满足 $1 \leqslant q = o(k)$ 的整数 $q = q_n$，令 $s = q^{1/3}$，并定义序列 $a_n(s) = (s+2)^3 - s^3, n \in \mathbf{N}$。则，

$$\Delta_{a_n(s)} \mathcal{F}^{\tilde{P}}_{\mathrm{ann},\lambda}(q) = (1 + o(1)) \Delta_{a_n(s)} \mathcal{F}^{P}_{\mathrm{ann},\lambda}(q), \tag{A.12}$$

对所有此类 q，上述结论一致成立。

证明。 首先注意，对于这两个模型，

$$\mathcal{F}_{\mathrm{ann},\lambda}(q) = -\lambda q - \log \mathbb{P}(\langle X, X' \rangle = q). \tag{A.13}$$

另请注意，对于这两个模型，若 X, X' 对应于向量 v, v', $\langle X, X' \rangle = S^3$, 其中 $S = \langle v, v' \rangle$。为证明定理，我们将首先证明以下结论，其中 $q = o(k)$,

$$\log \mathbb{P}_P(S^3 = q) = (1 + o(1)) \log \mathbb{P}(S^3 = q). \tag{A.14}$$

令 $L := \|v\|_0$。在独立同分布的 Rad 分布(k/n) 先验分布下， $L \sim$ 二项分布$(n, k/n)$ 和 $\mathbb{E}L = k$。因此，根据 Chernoff 不等式，存在 $c_0 > 0$，使得

$$\mathbb{P}\big(L \notin [k/2, 2k]\big) \leqslant 2e^{-c_0 k}.$$

由于 v, v' 相互独立,

$$\mathbb{P}(E_2^c) \leqslant 2\mathbb{P}\big(L \notin [k/2, 2k]\big) \leqslant 4e^{-c_0 k} = \exp\big(-\Omega(k)\big) = \exp\big(-\Omega(n^\beta)\big). \tag{A.15}$$

记 $X_i := v_i v_i' \in \{-1, 0, 1\}, i = 1, 2, \ldots, n$。那么 $S = \sum_{i=1}^n X_i$ 和

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) =: p, \qquad \mathbb{P}(X_i = 0) = 1 - 2p,$$

其中 $np = \Theta(k^2/n) = \Theta(n^{2\beta - 1}) \to 0$，因为 $\beta < 1/2$。

令 $s = q^{1/3}$。考虑恰有 s个坐标满足 $X_i = 1$，而其余 $n-s$个满足$X_i = 0$的事件。该事件意味着 $S = s$，因此

$$\mathbb{P}(S = s) \;\geqslant\; \binom{n}{s} p^s (1-2p)^{n-s}.$$

由于 $np \to 0$, 对于所有足够大的 n，我们都有 $(1-2p)^{n-s} \geqslant 1/2$。此外， $\binom{n}{s} \geqslant (n/s)^s$. 因此，对于所有足够大的 n，

$$\mathbb{P}(S = s) \geqslant \frac{1}{2}\Big(\frac{n}{s}\Big)^s p^s = \frac{1}{2}\Big(\frac{np}{s}\Big)^s.$$

由于 $np = n^{-(1-2\beta) + o(1)}$，可得

$$\mathbb{P}(S = s) \geqslant \exp\big(-C(s \log n + s \log s)\big) \tag{A.16}$$

对于某个常数 $C < \infty$。根据假设 $q = o(k)$，可得 s 对数 $n + s$对数 $s = o(k)$，从而我们推得

$$\mathbb{P}(S = s) \geqslant \exp(-o(k)). \tag{A.17}$$

对于上界，若 $S = s$，则必然至少有 s 个坐标满足 $X_i = 1$，因此

$$\mathbb{P}(S = s) \leqslant \sum_{j=s}^{n} \binom{n}{j} p^j (1-p)^{n-j} \leqslant \sum_{j=s}^{\infty} \frac{(np)^j}{j!} = O((np)^s),$$

这里用到了 $\binom{n}{j} \leqslant n^j/j!$ $1 - p \leqslant 1$，以及 $np \to 0.$。因此，

$$\log \mathbb{P}(S = s_n) \to -\infty. \tag{A.18}$$

现在令 $A_n := \{S^3 = s^3\} = \{S = s\}$。则

$$\mathbb{P}_P(A_n) = \frac{\mathbb{P}(A_n \cap E_2)}{\mathbb{P}(E_2)} = \mathbb{P}(A_n) \cdot \frac{1 - \mathbb{P}(A_n \cap E_2^c)/\mathbb{P}(A_n)}{1 - \mathbb{P}(E_2^c)}.$$

利用（A.15）和（A.16），

$$0 \leqslant \frac{\mathbb{P}(A_n \cap E_2^c)}{\mathbb{P}(A_n)} \leqslant \frac{\mathbb{P}(E_2^c)}{\mathbb{P}(A_n)} \leqslant \exp(-\Omega(k) + o(k)) = n^{-\omega(1)},$$

并且 $\mathbb{P}(E_2) = 1 - \mathbb{P}(E_2^c) = 1 - o(1)$。因此

$$\mathbb{P}_P(A_n) = \mathbb{P}(A_n)\,(1 + n^{-\omega(1)}). \tag{A.19}$$

取对数可得，利用 $|\log(x+1) - x| \leqslant x^2/2$ 对于 $x \to 0$,

$$\log \mathbb{P}_P(A_n) = \log \mathbb{P}(A_n) + n^{-\omega(1)}. \tag{A.20}$$

最后，由（A.18），

$$\log \mathbb{P}_P(A_n) = (1 + n^{-\omega(1)}) \log \mathbb{P}(A_n), \tag{A.21}$$

这正是（A.14）。为了得到离散导数的结论，在 s 和 $s + 2$处分别应用方程（A.21）。由于对任意 GAM,

$$\mathcal{F}_{\mathrm{ann},\lambda}(s^3) = -\lambda s^3 - \log \mathbb{P}(S = s),$$

利用 $s \geqslant 1$，我们得到,

$$\Delta_{a_n(s)} \mathcal{F}^{\tilde{P}}_{\mathrm{ann},\lambda}(s^3) = \frac{\mathcal{F}^{\tilde{P}}_{\mathrm{ann},\lambda}((s+2)^3) - \mathcal{F}^{\tilde{P}}_{\mathrm{ann},\lambda}(s^3)}{(s+2)^3 - s^3} \tag{A.22}$$

$$= -\lambda - \frac{\log \mathbb{P}_{\tilde{P}}(S = s+2) - \log \mathbb{P}_{\tilde{P}}(S = s)}{(s+2)^3 - s^3} \tag{A.23}$$

$$= -\lambda - (1 + n^{-\omega(1)}) \frac{\log \mathbb{P}(S = s+2) - \log \mathbb{P}(S = s)}{(s+2)^3 - s^3} \tag{A.24}$$

$$= (1 + o(1)) \Delta_{a_n(s)} \mathcal{F}^{P}_{\mathrm{ann},\lambda}(s^3) + \lambda n^{-\omega(1)}. \tag{A.25}$$

接下来，结合这样一个事实并利用 $n^{-C} \leqslant \lambda \leqslant n^C$，

$$-\Delta_{a_n} \log \mathbb{P}(\langle X, X' \rangle = q) \geqslant n^{C'}$$

对于某个 $C' > 0$ 且对所有 $q = o(k)$，由引理 10.5可得

$$\left|\Delta_{a_n(s)} \mathcal{F}^{P}_{\mathrm{ann},\lambda}\right| \geqslant n^{-(C+C')}. \tag{A.26}$$

公式（A.26）与（A.25）共同推出（A.12），证明至此完成。

□

A.2.3 “容易”区域中的小低次数 MMSE

本节中，我们证明，当 $\lambda = \tilde{\omega}(1)$，截断先验分布的低阶 MMSE 确实会变小，这是证明等价性的关键一步。

我们定义

$$\mathrm{MMSE}^{\leqslant D}(v) := \inf_{\widehat{v}} \mathbb{E}\|v - \widehat{v}(Y)\|_2^2, \qquad \mathrm{MMSE}^{\leqslant D}(X) := \inf_{\widehat{X}} \mathbb{E}\|X - \widehat{X}(Y)\|_F^2.$$

引理 A.5 （将张量估计化为向量估计）。 存在一个正数常量 $C > 0$ ，使得对于所有 $D > 0$,

$$\mathrm{MMSE}^{\leqslant D}(X) \leqslant C\, k^2 \mathrm{MMSE}^{\leqslant D}(v).$$

证明。令

$$\mathcal{C} := \{x \in \mathbb{R}^n : \ \|x\|_\infty \leqslant 1,\ \|x\|_0 \leqslant 2k\}.$$

由于真实的 向量 v 几乎处处属于$\mathcal{C}$ ，因此对于任意估计量 $\widehat{v}(Y)$ ，我们可以将其投影到$\mathcal{C}$上:

$$\widetilde{v}(Y) := \Pi_{\mathcal{C}}(\widehat{v}(Y)).$$

由于 C 是闭集， $\|v - \widetilde{v}\|_2 \leqslant \|v - \widehat{v}\|_2$ 几乎处处。此外， $\|v\|_2 \leqslant \sqrt{2k}$, $\|\widetilde{v}\|_2 \leqslant \sqrt{2k}$. 现在定义$\widetilde{X} := \widetilde{v}^{\otimes 3}$. 利用望远镜恒等式

$$v^{\otimes 3} - \widetilde{v}^{\otimes 3} = \sum_{j=1}^{r} v^{\otimes(j-1)} \otimes (v - \widetilde{v}) \otimes \widetilde{v}^{\otimes(3-j)},$$

结合恒等式 $\|v^{\otimes m}\|_F = \|v\|_2^m$，我们得到

$$\|X - \widetilde{X}\|_F \leqslant \sum_{j=1}^{3} \|v\|_2^{j-1} \|v - \widetilde{v}\|_2 \|\widetilde{v}\|_2^{3-j} \leqslant 6k \|v - \widetilde{v}\|_2.$$

平方并取期望可得

$$\mathbb{E}\|X - \widetilde{X}\|_F^2 \leqslant 9k^2 \mathbb{E}\|v - \widetilde{v}\|_2^2 \leqslant 9k^2 \mathbb{E}\|v - \widehat{v}\|_2^2.$$

最后，对所有估计量 $\widehat{v}$ 取下确界，得到所需结果。 □

引理 A.6（ v 的一个对角奇次幂估计量）。令 $d \geqslant 1$ 为一个奇整数，并定义

$$\widehat{v}_i := \lambda^{-d} Y^d_{i,\dots,i},\ i = 1, \dots, n. \quad \textit{Then,}\ \ \mathbb{E}\|v - \widehat{v}\|_2^2 \leqslant \frac{8k\, d^3}{\lambda^2} + n\Big(\frac{2d}{\lambda^2}\Big)^d \tag{A.27}$$

在 $2d^2 \leqslant \lambda^2$. 的条件下

证明。固定 $i \in \{1, \dots, n\}$。

情形 *1*： $v_i = 0$。 于是 $Y_{i,\dots,i} = Z_{i,\dots,i}$, 因此 $\widehat{v}_i = \lambda^{-d} Z^d_{i,\dots,i}$, 故

$$\mathbb{E}[\widehat{v}_i^2 \mid v_i = 0] = \lambda^{-2d} \mathbb{E}[Z^{2d}_{i,\dots,i}] = \lambda^{-2d} (2d-1)!!.$$

由于 $(2d-1)!! = 1 \cdot 3 \cdot 5 \cdots (2d-1) \leqslant (2d)^d$, 我们得到

$$\mathbb{E}[\widehat{v}_i^2 \mid v_i = 0] \leqslant \Big(\frac{2d}{\lambda^2}\Big)^d.$$

情况 *2*： $v_i = 1$。现在$Y_{i,\ \dots,\ i} = \lambda + Z_{i,\ \dots,\ i}$, 因此

$$\widehat{v}_i - 1 = \lambda^{-d}(\lambda + Z_{i,\dots,i})^d - 1 = \sum_{j=1}^{d} \binom{d}{j} \lambda^{-j} Z^j_{i,\dots,i}.$$

因此，利用柯西–施瓦茨不等式的形式 $\left(\sum_{j=1}^{d} a_j\right)^2 \leqslant d\sum_{j=1}^{d} a_j^2$，我们得到

$$\mathbb{E}[(\widehat{v}_i - 1)^2 \mid v_i = 1] \leqslant d\sum_{j=1}^{d} \binom{d}{j}^2 \lambda^{-2j}\mathbb{E}[Z_{i,\dots,i}^{2j}].$$

使用

$$\binom{d}{j} \leqslant \frac{d^j}{j!}, \qquad \mathbb{E}[Z_{i,\dots,i}^{2j}] = (2j-1)!! \leqslant 2^j j!,$$

我们得到

$$\mathbb{E}[(\widehat{v}_i - 1)^2 \mid v_i = 1] \leqslant d\sum_{j=1}^{d} \frac{d^{2j}}{(j!)^2}\lambda^{-2j}\cdot 2^j j! = d\sum_{j=1}^{d} \frac{(2d^2/\lambda^2)^j}{j!} \leqslant d\big(e^{2d^2/\lambda^2} - 1\big).$$

在假设 $2d^2 \leqslant \lambda^2$ 下，我们有 $2d^2/\lambda^2 \leqslant 1,$，因此使用 $e^x - 1 \leqslant 2x,\ \ x \in [0,\ \ 1]$ 可得

$$\mathbb{E}[(\widehat{v}_i - 1)^2 \mid v_i = 1] \leqslant \frac{4d^3}{\lambda^2}.$$

情形 *3*： $v_i = -1$。由于 d 为奇数且 $Z_{i,\ \dots\dots,\ i} \overset{d}{=} -Z_{i,\ \dots\dots,\ i}$，

$$\mathbb{E}[(\widehat{v}_i + 1)^2 \mid v_i = -1] = \mathbb{E}[(\widehat{v}_i - 1)^2 \mid v_i = 1] \leqslant \frac{4d^3}{\lambda^2}.$$

现在对所有坐标求和。由于 $\|v\|_0 \leqslant 2k$ 几乎处处成立，活跃坐标至多有 $2k$ 个，而坐标总数至多为 n 个。因此

$$\mathbb{E}\|v - \widehat{v}\|_2^2 \leqslant 2k\cdot\frac{4d^3}{\lambda^2} + n\Big(\frac{2d}{\lambda^2}\Big)^d,$$

这正是（A.27）。 □

推论 A.7。 假设 $\lambda \geqslant (\text{对数}\, n)^2.$ 令 D 为任意值，且满足 $D = \Theta(\text{对数}\ n)$。则，

$$\mathrm{MMSE}^{\leqslant D}(X) = o(k^3).$$

证明。 由于 $D \leqslant$ 对数 n 且 $\lambda \geqslant (\text{对数}\, n)^2$，对于所有充分大的 n，有 $2D^2 \leqslant 2(\text{对数}\ n)^2 \leqslant \lambda^2$，因此引理 A.6适用。于是

$$\mathbb{E}\|v - \widehat{v}\|_2^2 \leqslant \frac{8k\,D^3}{\lambda^2} + n\Big(\frac{2D}{\lambda^2}\Big)^D.$$

对于第一项，

$$\frac{8k\,D^3}{\lambda^2} \leqslant 8k\frac{(\log n)^3}{(\log n)^4} = \frac{8k}{\log n} = o(k).$$

对于第二项，由于 $D = \Theta(\text{对数}\, n)$ 且 $\lambda^2 \geqslant (\text{对数}\ n)^4$，

$$n\Big(\frac{2D}{\lambda^2}\Big)^D \leqslant n\Big(\frac{2\log n}{(\log n)^4}\Big)^D = n\Big(\frac{2}{(\log n)^3}\Big)^D.$$

取对数并利用 $D = \Theta(\text{对数}\, n)$，

$$\log\left[n\Big(\frac{2}{(\log n)^3}\Big)^D\right] = \log n + D\big(\log 2 - 3\log\log n\big) \to -\infty.$$

因此 $n(2D/\lambda^2)^D = o(1),$，从而它当然是 $o(k)$。这就证明了

$$\mathbb{E}\|v - \widehat{v}\|_2^2 = o(k),$$

因此，最小均方误差 $\leqslant^D(v) = o(k)$。最小均方误差$\leqslant^D(X) = o(k^3)$ 结论随后由推论 A.7得出。 □

A.2.4 综合起来

定理 A.8（截断模型的退火 FP 单调性与低阶 MMSE 一致性）。固定 $C > 0$ 并假设 $n^{-C} \leqslant \lambda \leqslant n^{C}$。存在 λ_1，$\lambda_2 = \tilde{\Theta}(1)$ 满足 $\lambda_1 < \lambda_2$，使得以下结论成立。

若 $\lambda \leqslant \lambda_1$，则

1. 截断模型的退火 *Franz–Parisi* 势 $F^{P}_{\mathrm{ann},\ \lambda}(q)$ 在区间 $q = s^3 \in [0, \Theta((\text{对数}\, n)^3)]$.上单调递减

2. 对任意 $D = \Theta(\text{对数}\, n)$，低阶 *MMSE* 渐近平凡，即

$$\mathrm{MMSE}_{P}^{\leqslant D}(\lambda) = (1 - o(1))\mathrm{MMSE}_{P}^{\mathrm{trivial}}$$

若 $\lambda \geqslant \lambda_2$，则

1. 截断模型的退火 *Franz–Parisi* 势 $F^{P}_{\mathrm{ann},\ \lambda}(q)$ 在区间 $q = s^3 \in [0, n^{o(1)}]$.上单调递增

2. 低次　　次 *MMSE* 对于 $D = \Theta(\text{对数}\, n)$ 渐近地低于平凡值，即　　*ly*

$$\mathrm{MMSE}_{P}^{\leqslant D}(\lambda) = o(\mathrm{MMSE}_{P}^{\mathrm{trivial}}).$$

注 A.9。请注意，这项任务的两种不同预测是一致的。当 λ 小于算法阈值时，由于退火 FP 递增，该问题属于“物理困难”情形；同时，由于低阶 MMSE 是平凡的，该问题也属于低次困难情形。另一方面，如果 λ 大于算法阈值，由于退火 FP 递减，该问题属于“物理容易”情形；同时，由于低阶 MMSE 优于平凡 MSE，该问题也属于低次困难情形。

˜证明。 我们使用符号 P和 P分别表示未截断先验和截断先验，其定义见第A.2.1节开头。

˜首先，根据定理 3.2 对于未截断先验，存在 $\lambda_1 = \Theta(1)$ 使得若 $\lambda \leqslant \lambda_1$，则对某个 $D = \Theta(\text{对数}\, n)$,

$$\mathrm{MMSE}_{\tilde{P}}^{\leqslant D}(\lambda) = (1 + o(1))\mathrm{MMSE}^{\mathrm{trivial}} = (1 + o(1))k^r.$$

根据命题 A.3,

$$\left|\mathrm{MMSE}_{P}^{\leqslant D}(\lambda) - \mathrm{MMSE}_{\tilde{P}}^{\leqslant D}(\lambda)\right| \leqslant e^{-\Theta(k)}$$

因此

$$\mathrm{MMSE}_{P}^{\leqslant D}(\lambda) = (1 + o(1))\mathrm{MMSE}^{\mathrm{trivial}} = (1 + o(1))k^r.$$

此外，对任意 $1 \leqslant q = s^3 = n^{o(1)}$，以及序列 $a_n(s) = (s+2)^3 - s^3$，$n \in \mathbf{N}$ 由引理10.5可得,

$$\Delta_{a_n(s)}\mathcal{F}^{P}_{\mathrm{ann},\lambda}(q) = -\lambda + \Theta\left(\frac{\log n}{q^{2/3}}\right), \tag{A.28}$$

其中 Θ 包含普适常数（与 q无关）。特别地，若有必要，进一步缩小 λ_1，并假设 $\lambda_1 = o(1/\text{对数}\ n)$, 则对任意 $\lambda \leqslant \lambda_1$，退火 FP 势在区间 $q = s^3 \in [0,\ \Theta((\text{对数}\, n)^3)]$. 现在由定理 A.4可得,

$$\Delta_{a_n(s)}\mathcal{F}^{\tilde{P}}_{\mathrm{ann},\lambda}(q) = (1 + o(1))\Delta_{a_n(s)}\mathcal{F}^{P}_{\mathrm{ann},\lambda}(q), \tag{A.29}$$

因此,　　截断先验的退火 FP 势也具有相同的单调行为。

˜现在，对于某些 $\lambda_2 = \Theta(1)$，根据推论 A.7 可知，若 $\lambda \geqslant \lambda_2$ 对于任意 $D = \Theta(\text{对数}\, n)$ 以及满足以下条件的截断先验分布,

$$\mathrm{MMSE}_{P}^{\leqslant D}(\lambda) = o(\mathrm{MMSE}^{\mathrm{trivial}}).$$

此外，固定任意 $1 \leqslant q = s^3 = n^{o(1)}$，以及序列 $a_n(s) = (s+2)^3 - s^3$，$n \in \mathbf{N}$。由（A.28）并假设 $\lambda_2 = \omega(\text{对数}\, n)$，可得，对于 $\lambda \geqslant \lambda_2$，未截断先验的退火 FP 势对所有 $1 \leqslant q = n^{o(1)}$. 现在根据定理 A.4，截断先验的退火 FP 势也具有相同的单调行为。

□

A.3 淬　火 FP 势在“低次多项式容易”相中并非单调递减

现在我们陈述并证明以下定理，证明即使在“低次”易区间 $\lambda = k^{\Omega(1)}$（此时低阶 MMSE 是平凡的），淬火 FP 势仍然是非单调的。特别是，淬火 FP 势的单调性无法捕捉该模型中的低阶 MMSE 相变。值得注意的是，正如我们在上一节所讨论的，退火 FP 势的单调性确实能够准确反映这一相变。

定理 A.10. 存在常数 $c, C, C' > 0$，使得下述结论成立。若 $\beta > 0$ 是一个足够小的常数，$k = n^{\beta + o(1)}$，则对任意 $q \geqslant 1$,

$$\mathcal{F}_\lambda(q) - \mathcal{F}_\lambda(0) \geqslant -\lambda^2 q - Ck\log(n/k) + c\lambda k q^{1/3}/\sqrt{\log n}.$$

特别地，若 $\lambda = k^{\varepsilon + o(1)}$ 对任意 $\varepsilon \in (0,1)$ 均成立（即对 $\lambda \geqslant n^{\Omega(1)}\lambda_{\mathrm{ALG}}$的取值均成立），那么只要 C(对数 $n)^3 \leqslant q \leqslant$ (对数 $n)^d$ 对某个 $d > 3$成立，就有

$$\mathcal{F}_\lambda(q) - \mathcal{F}_\lambda(0) \geqslant k^{1+\Omega(1)} > 0,$$

因此，对所有 $d > 3$，淬火 *FP* 势 $F_\lambda(q)$ 不是关于 $q \in [0, (\text{对数 } n)^d]$的单调递减函数 .

A.3.1 关键引理

为证明这一点，我们首先需要几个引理。随机曲率至关重要 e

$$\Gamma(q', m) := \max_{v' \in S_{n,m} : \langle v, v' \rangle = q'} \langle \mathrm{vec}(v'^{\otimes 3}), Z \rangle, q' \in \mathbb{R}, m \in [n]$$

其中 S_n $_m$是 $\{-1$中的 m-稀疏向量集合，0，$1\}^n$ 并且照常 Z 具有独立同分布的 $N(0,\ 1)$ 个条目 s.

引理 A.11。假设 $\beta > 0$ 是一个足够小的常数。那么存在一个常数 $C > 0$，使得对于任意 $1 \leqslant q' = o(k)$、$k/2 \leqslant m \leqslant 2k$ 以及任意增长序列 A_n，以至少 $1 - Ce^{-A_n/2}$, 的概率有

$$\Gamma(q', m) \leqslant \sqrt{m^3}\sqrt{2\log\left(k\binom{k}{q'}\binom{n-k}{m-q'}\right) - \log(m\log(n/m)) + A_n}. \tag{A.30}$$

特别地，当 $C > 0$ 足够大时，若 $A_n \geqslant C$ 对数 n，则对于任意 $1 \leqslant q' = o(k)$，以概率$1 - e^{-\Theta(A_n)}$，(A.30) 对所有 $k/2 \leqslant m \leqslant 2k$同时成立。

此外，对于任意 $c > 0$，若 $\alpha > 0$ 足够小，则对于所有满足 $k/2 \leqslant m \leqslant 2k$, 的 m

$$\mathbb{E}[\Gamma(q', m) 1(\Gamma(q', m) \geqslant n^c)] = n^{-\Omega(1)}.$$

证明。对于任意 $x \in \{-1,\ 0,\ 1\}^n$（其为 m-稀疏），注意到 $\langle Z,\ \mathrm{vec}(x^{\otimes 3})\rangle$ 是一个均值为零、方差为 m^3 的高斯分布。

注意，满足 $\langle x,\ v\rangle = q'$，的 $x \in \{-1,\ 0,\ 1\}^n$（其为 m-稀疏）的数量为

$$\sum_{b \geqslant q'} \binom{b}{q'}\binom{k}{b}\binom{n-k}{m-b} \leqslant k\binom{k}{q'}\binom{n-k}{m-q'}.$$

我们首先利用恒等式化简被加数

$$\binom{k}{b}\binom{b}{q'} = \binom{k}{q'}\binom{k-q'}{b-q'},$$

因此，记为 $t = b - q'$,

$$\sum_{b \geqslant q'} \binom{b}{q'}\binom{k}{b}\binom{n-k}{m-b} = \binom{k}{q'}\sum_{t \geqslant 0}\binom{k-q'}{t}\binom{n-k}{(m-q')-t}.$$

根据范德蒙德恒等式，

$$\sum_{t\geqslant 0}\binom{k-q'}{t}\binom{n-k}{(m-q')-t}=\binom{(k-q')+(n-k)}{m-q'}=\binom{n-q'}{m-q'},$$

这就证明了精确等式

$$\sum_{b\geqslant q'}\binom{b}{q'}\binom{k}{b}\binom{n-k}{m-b}=\binom{k}{q'}\binom{n-q'}{m-q'}.$$

对于不等式，只需证明

$$\binom{n-q'}{m-q'}\leqslant k\binom{n-k}{m-q'}.$$

令 $r:=m-q'$。由于 $m\leqslant 2k$ 和 $q'\geqslant 1$，我们有 $0\leqslant r\leqslant 2k$。利用二项式系数的乘积表示，

$$\frac{\binom{n-q'}{r}}{\binom{n-k}{r}}=\prod_{i=0}^{r-1}\frac{n-q'-i}{n-k-i}=\prod_{i=0}^{r-1}\Big(1+\frac{k-q'}{n-k-i}\Big)\ \leqslant\ \exp\Big(\sum_{i=0}^{r-1}\frac{k-q'}{n-k-i}\Big).$$

由于 $r\leqslant 2k$ 和 $i\leqslant r-1\leqslant 2k-1$，我们有 $n-k-i\geqslant n-3k$。因此

$$\sum_{i=0}^{r-1}\frac{k-q'}{n-k-i}\leqslant\frac{r(k-q')}{n-3k}\leqslant\frac{2k^2}{n-3k}.$$

假设 $k\leqslant n^{\beta}$ 且满足 $\beta<\frac{1}{2}$。则 $k^2/n=n^{2\beta-1}\to 0,$，并且 $3k=o(n)$，因此对于所有充分大的 n，我们有 $\frac{2k^2}{n-3k}\leqslant$ 对数 2。因此，对于所有足够大的 n，

$$\frac{\binom{n-q'}{r}}{\binom{n-k}{r}}\leqslant e^{\log 2}=2\leqslant k,$$

这意味着 $\binom{n-q'}{r}\leqslant k\binom{n-k}{r}$，从而完成证明。

因此，根据并集界和米尔斯比率界，存在某个 x是 m-稀疏且满足 $\langle x、\ v\rangle=q'$ 的情况下，（A.30）不成立的概率至多为

$$\begin{aligned}&\frac{1+o(1)}{\sqrt{2\pi}}\frac{\exp\left(\frac{1}{2}\log\left(m\log(n/m)\right)-\frac{1}{2}A_n\right)}{\sqrt{2\log\left(k\binom{k}{q'}\binom{n-k}{m-q'}\right)-\log\left(m\log(n/m)\right)+A_n}}\\&=\frac{1+o(1)}{\sqrt{2\pi}}\sqrt{\frac{m\log(n/m)}{2\log\left(k\binom{k}{q'}\binom{n-k}{m-q'}\right)-\log\left(m\log(m/s)\right)+A_n}}\exp\left(-\frac{1}{2}A_n\right)\\&=O\left(\exp\left(-\frac{1}{2}A_n\right)\right).\end{aligned}$$

对于倒数第二个等式，我们利用了由于 $q'=o(k)$，从而 $q'=o(m)$，有

$$\log\left(\binom{k}{q'}\binom{n-k}{m-q'}\right)\geqslant\log\binom{n-k}{m-q'}\geqslant(1-o(1))m\log(n/m)$$

这就证明了（A.30）。

固定 $q'=o(k)$ 并令

$$\mathcal{S}:=\{m\in\mathbb{Z}:\ k/2\leqslant m\leqslant 2k\},\qquad|\mathcal{S}|\leqslant 2k.$$

对每个固定的 $m \in \mathcal{S}$，(A.30) 得出

$$\Pr\big((\text{A.30}) \text{ fails for this } s\big) \leqslant Ce^{-A_n/2}.$$

对 $m \in \mathcal{S}$ 应用并集界，得到

$$\Pr\Big(\exists m \in \mathcal{S}: \ (\text{A.30}) \text{ fails}\Big) \leqslant \sum_{m\in\mathcal{S}} Ce^{-A_n/2} \leqslant 2k\, Ce^{-A_n/2}.$$

特别地，若 $A_n = \omega(\text{对数}\, n)$，且 $k = n^\beta$ 对某个固定的 $\beta > 0$，则 $2kCe^{-A_n/2} = \exp(-\Theta(A_n))$，因此（A.30）对所有 $m \in [k/2, 2k]$ 同时以概率 $1 - e^{-\Theta(A_n)}$成立。

固定任意 $c > 0$，并取足够小的 $\beta > 0$，使得 $2\beta < c$。对于 $m \in [k/2, 2k]$，定义阈值

$$B_{q',m} := \sqrt{m^3}\sqrt{2\log\Big(k\binom{k}{q'}\binom{n-k}{m-q'}\Big) - \log\big(m\log(p/m) + A_n\big)}.$$

利用 $\log\binom{k}{q'} \leqslant q'\log(ek/q') \leqslant k\log(ek)$（因为 $q' = o(k)$）、$\log\binom{n-k}{m-q'} \leqslant m\log(en/m)$（因为 $m \asymp k$）以及 $m \asymp k$，可得存在某个常数 $C > 0$，使得

$$\log\Big(k\binom{k}{q'}\binom{n-k}{m-q'}\Big) \leqslant Ck\log n;$$

因此，对于某个绝对常数 $C' > 0$,

$$B_{q',m} \leqslant C'k^{3/2}\sqrt{k\log n} = C'k^2\sqrt{\log n}.$$

由于 $k = n^\beta$ 和 $2\beta < c$，对于所有充分大的 n，我们有 $B_{q',\, m} \leqslant \frac{1}{2}n^c$，且关于 $m \in [k/2, 2k]$均匀成立。因此，对于所有这样的 n,

$$\Pr\big(\Gamma(q', m) \geqslant n^c\big) \leqslant \Pr\big(\Gamma(q', m) > B_{q',m}\big) \leqslant Ce^{-A_n/2}.$$

此外，$\Gamma(q',\ m)$ 是$k\binom{k}{q'}\binom{n-k}{m-q'}$ 个中心化次高斯变量的最大值，其代理方差为 $\leqslant cm^3$，其中某个 $c > 0$绝对常数（参见（A.30）的证明）。根据次高斯族的标准最大不等式（例如 [Ver18，命题 2.5。2]），存在常数 $C,\ C' > 0$，使得

$$\mathbb{E}[\Gamma(q', m)^2] \leqslant Cm^3\log\Big(k\binom{k}{q'}\binom{n-k}{m-q'}\Big) \leqslant C'k^3 \cdot (k\log n) = C'k^4\log n.$$

~应用柯西–施瓦茨不等式可得，对于某个绝对常数 $C > 0$,

$$\mathbb{E}\big[\Gamma(q', m)\mathbf{1}\{\Gamma(q', m) \geqslant n^c\}\big] \leqslant \big(\mathbb{E}[\Gamma(q', m)^2]\big)^{1/2}\Pr(\Gamma(q', m) \geqslant n^c)^{1/2} \leqslant \widetilde{C}k^2\sqrt{\log n}\, e^{-A_n/4}.$$

取例如 $A_n = (\log n)^2$（或任何增长足够快的 $A_n \to \infty$）可得，对于所有$m \in [k/2, 2k]$, 均匀成立的 $k^2\sqrt{\log n}\, e^{-A_n/4} = n^{-\Omega(1)},$，从而完成证明。□

我们还可以证明以下结论。

引理 A.12。假设 $0 < \beta < 1/5$。那么，对于任意 $k/2 \leqslant m \leqslant 2k$ 以及任意递增序列 A_n，以概率 $1 - n^{-\Omega(1)}$,

$$\Gamma(0, m) \geqslant \sqrt{m^3}\sqrt{2\log\binom{n}{m} - \log(m\log(n/m)) - A_n}.$$

证明。对 [CSZ24, 命题 9.4] 应用于 $t = 3$、稀疏度 m 和 $\ell = 1/2$，（当 $\beta < 1/5$ 时），并直接检验二阶矩方法得出的高概率保证 $1 - n^{-\Omega(1)}$，即可得到结论（具体而言，在 [CSZ24, 引理 9.7] 中，所有 $o(1)$ 项都可以直接验证为 $n^{-\Omega(1)}$）。需要指出的是，尽管形式上 [CSZ24] 证明了 [CSZ24, 命题 9.4] 仅适用于整数 ℓ，但对于 ℓ，证明经过相应调整后同样适用于非整数情形。此外，在 $\ell = 1/2$, $\Gamma_{[0,\ell]}$ 的情形下，[CSZ24, 命题 9.4] 等于 $\Gamma(0,\ m)/\sqrt{m^3}$ 用我们的符号表示即得本引理。□

基于我们的先验分布，令 $p(m,\ q') = \mathrm{P}_v(\langle v,\ v'\rangle = q',\ \|v'\|_0 = m)$.

引理 A.13. 成立

$$\min_{k/2\leqslant m\leqslant 2k} p(m,0) \geqslant e^{-\Theta(k)}.$$

证明。 固定 $m \in [k/2, 2k]$ 并考虑事件

$$E_m := \{\|v'\|_0 = m\} \cap \{\mathrm{supp}(v') \cap \mathrm{supp}(v) = \varnothing\}.$$

在 E_m 上，我们有 $\langle v,\ v'\rangle = 0$，因此 $p(m,\ 0) \geqslant \mathrm{P}(E_m)$. 在给定 v'的条件下，$\{\|v'\|_0 = m\}$ 的支撑集是某个集合 $S \subset [n]$，且满足 $|S| = m$。由各坐标之间的独立性，

$$\mathbb{P}\big(\mathrm{supp}(v') \cap \mathrm{supp}(v) = \varnothing \big|\ \|v'\|_0 = m, \mathrm{supp}(v') = S\big) = (1-k/n)^m.$$

因此

$$\mathbb{P}(E_m) = \mathbb{P}(\|v'\|_0 = m)(1-k/n)^m.$$

我们有 $\|v'\|_0 \sim$ 二项分布$(n,\ k/n)$ 其均值为 k。对于任意 $m \in [k/2, 2k]$, 二项分布的标准局部下界（例如参见 [Ver18，命题 2.1。2]）给出 $\mathrm{P}(\|v'\|_0 = m) \geqslant \exp(-C_1 k)$ 其中某个绝对常数为 $C_1 > 0$ 且对所有足够大的 n均成立。

由于 $m \leqslant 2k$ 且当 x较小时 $\log(1-x) \geqslant -x - x^2$，

$$(1-k/n)^m = \exp\big(m\log(1-k/n)\big) \geqslant \exp\Big(-\frac{km}{n} - O\Big(\frac{k^2 m}{n^2}\Big)\Big) \geqslant \exp\Big(-O\Big(\frac{k^2}{n}\Big)\Big).$$

由于 $\beta < 1/2,$，我们有 $k^2/n = n^{2\beta-1+o(1)} = o(1),$，因此

$$(1-k/n)^m = \exp(-o(k)).$$

结合这些界，

$$p(m,0) \geqslant \mathbb{P}(E_m) \geqslant \exp(-C_1 k)\cdot\exp(-o(k)) = \exp(-\Theta(k)),$$

对所有 $m \in [k/2, 2k]$ 均一致成立。对 m 取最小值即可完成证明。 □

A.3.2 定理的证明A.10

根据淬火 FP 势的定义，对于任意 $q = (q')^3 > 0$，我们有：

$$\mathcal{F}_\lambda(q) = -\mathbb{E}_{v,Z}\left[\log \mathbb{E}_{v'}\left[\mathbf{1}_{\{\langle v,v'\rangle = q'\}} \exp\left(-\frac{1}{2}\|Y - \sqrt{\lambda}\mathrm{vec}((v')^{\otimes 3})\|_2^2\right)\right]\right]$$

其中 $Z \sim \mathcal{N}(0,\ I_N)$。展开范数平方，并去掉与 $q = (q')^3$ 无关的项后，自由能之差可简化为：

$$\mathcal{F}_\lambda(q) - \mathcal{F}_\lambda(0) = -\lambda^2 q - \mathbb{E}_{v,Z}\left[\log \frac{\mathbb{E}_{v'}\left[\mathbf{1}_{\{\langle v,v'\rangle = q'\}} e^{\lambda\langle \mathrm{vec}((v')^{\otimes 3}), Z\rangle}\right]}{\mathbb{E}_{v'}\left[\mathbf{1}_{\{\langle v,v'\rangle = 0\}} e^{\lambda\langle \mathrm{vec}((v')^{\otimes 3}), Z\rangle}\right]}\right].$$

为了界定这些配分函数的比值，我们按稀疏度级别$m = \|v'\|_0 \in [\lfloor k/2\rfloor, 2k]$对内层期望进行分层。对于分子，我们用 m 上的最大值和支撑集 $S_{n,\ m}$的大小来界定该和：

$$\sum_{m=\lfloor k/2\rfloor}^{2k} \mathbb{E}_{v'}\left[\mathbf{1}_{\{\langle v,v'\rangle = q', \|v'\|_0 = m\}} e^{\lambda\langle \mathrm{vec}((v')^{\otimes 3}), Z\rangle}\right] \leqslant \max_{m\in[\lfloor k/2\rfloor, 2k]}\left(e^{\lambda\Gamma(q',m)} p(m,q')\right)\cdot \sum_{m=\lfloor k/2\rfloor}^{2k} |S_{n,m}|.$$

由于稀疏支撑集的总数可宽松地上界为 $\sum_{m=\lfloor k/2\rfloor}^{2k} |S_{n,m}| \leqslant 2k\binom{n}{2k}2^{2k},$，其对数为 $O(k\log(n/k))$。

对于分母，我们用其最大项作为该和的下界：

$$\sum_{m=\lfloor k/2\rfloor}^{2k} \mathbb{E}_{v'}\left[\mathbf{1}_{\{\langle v,v'\rangle=0,\|v'\|_0=m\}} e^{\lambda\langle \mathrm{vec}((v')^{\otimes 3}),Z\rangle}\right] \geqslant \max_{m\in[\lfloor k/2\rfloor,2k]} \left(e^{\lambda\Gamma(0,m)}p(m,0)\right).$$

对这些界取比值并应用对数后，自由能差有如下界：

$$\mathcal{F}_\lambda(q)-\mathcal{F}_\lambda(0) \geqslant -\lambda^2 q - O\left(k\log\frac{n}{k}\right) - \mathbb{E}_Z\left[\max_{m\in[\lfloor k/2\rfloor,2k]} \lambda\big(\Gamma(q',m)-\Gamma(0,m)\big)\right] - \max_{m\in[\lfloor k/2\rfloor,2k]} \log\frac{1}{p(m,0)}.$$

根据引理 A.13，我们已建立 $\min_{k/2\leqslant m\leqslant 2k} p(m,0) \geqslant e^{-\Theta(k)}$,，这将被吸收到现有的$O(k$对数$(n/k))$ 项中。此外，以概率 $1-n^{-\Omega(1)}$，根据引理 A.11 和 A.12，最大值之间的偏差对所有 $k/2\leqslant m\leqslant 2k$ 均同时有界：

$$\Gamma(q',m)-\Gamma(0,m) \leqslant -\frac{m(q'-A_n-\log k)}{\Theta(\sqrt{\log n})}.$$

应用引理A. 11 来控制尾部期望，我们发现对于 $m>2(A_n+$对数 $k)$，并且由于 $q=(q')^3$ ，

$$\mathbb{E}_Z\left[\max_{m\in[\lfloor k/2\rfloor,2k]} \lambda\big(\Gamma(q,s)-\Gamma(0,s)\big)\right] \leqslant -\Theta\left(\frac{\lambda k q^{1/3}}{\sqrt{\log n}}\right)+o(1).$$

将该期望界代回自由能不等式可得，对于某些常数C， $c>0$:

$$\mathcal{F}_\lambda(q)-\mathcal{F}_\lambda(0) \geqslant -\lambda^2 q - Ck\log\frac{n}{k} + c\lambda\frac{kq^{1/3}}{\sqrt{\log n}}.$$

最后，对于足够大的 $C>0$，我们取 $A_n=C$对数 n。若对任意 $\varepsilon\in(0$设置 $\lambda=k^{\varepsilon+o(1)}$， $1)$并在 $q\geqslant(2C)^3$(对数$n)^3$处进行求值，则正项严格占优。因此，我们得到：

$$\mathcal{F}_\lambda(q)-\mathcal{F}_\lambda(0) \geqslant k^{1+\Omega(1)} > 0.$$

B 辅助引理的证明

在本节中，我们证明命题 6.3 和7.3，以及引理 A.2 和 6.8。

引理 *A.2* 的证明。 我们使用标准高斯尾界：对于 $Z\sim\mathcal{N}(0,\ 1)$ 以及 $t\geqslant 0$,

$$\mathbb{P}(|Z|\geqslant t) \leqslant 2e^{-t^2/2}. \tag{B.1}$$

回顾一下， $d_i=\lambda v_i+Z_i$其中 $Z_i\sim\mathcal{N}(0,\ 1)$独立同分布。如果 $v_i=0$，则 $d_i=Z_i$。事件 $\widehat{v}_i\neq v_i$为 $\{|d_i|\geqslant\tau\}$，因此根据并集界和（B.1），

$$\mathbb{P}\big(\exists i:\ v_i=0,\ \widehat{v}_i\neq 0\big) \leqslant \sum_{i:\,v_i=0}\mathbb{P}(|Z_i|\geqslant\tau) \leqslant n\cdot 2e^{-\tau^2/2}=2n^{-2}, \tag{B.2}$$

因为$\tau=\sqrt{6$对数 $n}$。

接下来，固定 $i\in S=\mathrm{supp}(v)$。则 $v_i\in\{\pm1\}$和 $d_i\sim\mathcal{N}(\pm\lambda,\ 1)$。事件 $\widehat{v}_i=0$为 $\{|d_i|<\tau\}$，因此

$$\mathbb{P}(|d_i|<\tau) \leqslant 2\exp\Big(-\frac{1}{2}(\lambda-\tau)_+^2\Big) \leqslant 2e^{-\tau^2/2},$$

其中，第一个不等式与您原始证明中使用的双尾界相同，第二个不等式使用了 $\lambda\geqslant 2\tau$，因此 $\lambda-\tau\geqslant\tau$。对 $i\in S$ 取并集界并利用 $|S|\leqslant 2k$，得到

$$\mathbb{P}\big(\exists i\in S:\ \widehat{v}_i=0\big) \leqslant |S|\cdot 2e^{-\tau^2/2} \leqslant 4k\,e^{-\tau^2/2}=4k\,n^{-3}. \tag{B.3}$$

现在，固定 $i \in S$。若 $v_i = 1$，则 $\widehat{v}_i = -1$ 蕴含 $d_i \leqslant -\tau$；若 $v_i = -1$，则 $\widehat{v}_i = 1$ 蕴含 $d_i \geqslant \tau$。因此，对于每个 $i \in S$，

$$\mathbb{P}(\widehat{v}_i = -v_i) \leqslant \mathbb{P}\big(\mathcal{N}(\lambda, 1) \leqslant -\tau\big) + \mathbb{P}\big(\mathcal{N}(-\lambda, 1) \geqslant \tau\big) \leqslant 2\exp\Big(-\frac{1}{2}(\lambda + \tau)^2\Big),$$

根据（B.1）。在 $\lambda \geqslant 2\tau$ 条件下，我们有 $\lambda + \tau \geqslant 3\tau$，因此

$$2\exp\Big(-\frac{1}{2}(\lambda + \tau)^2\Big) \leqslant 2\exp\Big(-\frac{1}{2}(3\tau)^2\Big) = 2e^{-27\log n} = 2n^{-27}.$$

对 $i \in S$ 使用并集界，并利用 $|S| \leqslant 2k$，可得

$$\mathbb{P}\big(\exists i \in S:\ \widehat{v}_i = -v_i\big) \leqslant |S| \cdot 2n^{-27} \leqslant 4k\, n^{-27}. \tag{B.4}$$

最后，

$$\{\widehat{v} \neq v\} \subseteq \big\{\exists i:\ v_i = 0,\ \widehat{v}_i \neq 0\big\} \ \cup\ \big\{\exists i \in S:\ \widehat{v}_i = 0\big\} \ \cup\ \big\{\exists i \in S:\ \widehat{v}_i = -v_i\big\}.$$

结合（B.2）、（B.3）、（B.4）中的界，可得（A.2）。 □

命题 *6.3* 的证明。令 $V_i = Z_i$，其中 $i \neq t$，并令 $V_t = aX + bY$。固定一个划分 $\pi \in \mathcal{P}([n])$，并令 B_π 为包含 π 的唯一块，其中包含 t。那么

$$\mathbb{E}\Big[\prod_{i \in B_\pi} V_i\Big] = \mathbb{E}\Big[(aX + bY)\prod_{i \in B_\pi \setminus \{t\}} Z_i\Big] = a\,\mathbb{E}\Big[X\prod_{i \in B_\pi \setminus \{t\}} Z_i\Big] + b\,\mathbb{E}\Big[Y\prod_{i \in B_\pi \setminus \{t\}} Z_i\Big],$$

根据期望的线性性。对于每个满足 $t \notin B$ 的其他块 $B \in \pi$，我们有 $\mathbb{E}[\prod_{i \in B} V_i] = \mathbb{E}[\prod_{i \in B} Z_i]$，它不依赖于 X 或 Y。

因此，对于这个固定的 π，

$$\prod_{B \in \pi} \mathbb{E}\Big[\prod_{i \in B} V_i\Big] = a\left(\mathbb{E}\Big[X\prod_{i \in B_\pi \setminus \{t\}} Z_i\Big]\prod_{B \in \pi:\, B \neq B_\pi} \mathbb{E}\Big[\prod_{i \in B} Z_i\Big]\right) + b\left(\mathbb{E}\Big[Y\prod_{i \in B_\pi \setminus \{t\}} Z_i\Big]\prod_{B \in \pi:\, B \neq B_\pi} \mathbb{E}\Big[\prod_{i \in B} Z_i\Big]\right).$$

乘以系数 $(|\pi| - 1)!\,(-1)^{|\pi|-1}$ 并对所有划分 π 求和，我们可以将标量 a 和 b（求和是有限的）提出求和号外，并利用命题 6.4，恰好得到所声称的关于第t个自变量的线性性。 □

引理 *6.8* 的证明。展开范数平方，

$$\mathrm{MMSE}_X^{\leqslant D} = \inf_{f_1, \ldots, f_n \in \mathbb{R}[Y]_{\leqslant D}} \sum_{i=1}^n \mathbb{E}\big[(f_i(Y) - X_i)^2\big] = \sum_{i=1}^n \inf_{f_i \in \mathbb{R}[Y]_{\leqslant D}} \mathbb{E}\big[(f_i(Y) - X_i)^2\big].$$

对于每个 i，展开平方：

$$\mathbb{E}[(f_i(Y) - X_i)^2] = \mathbb{E}[f_i(Y)^2] - 2\mathbb{E}[f_i(Y)X_i] + \mathbb{E}[X_i^2].$$

对于任意固定的 $f_i \neq 0$，考虑 $g_i = \alpha f_i$，对标量 $\alpha \in \mathbb{R}$ 取最小值：

$$\inf_{\alpha \in \mathbb{R}} \mathbb{E}[(\alpha f_i(Y) - X_i)^2] = \mathbb{E}[X_i^2] - \frac{\mathbb{E}[f_i(Y)X_i]^2}{\mathbb{E}[f_i(Y)^2]}.$$

事实上，关于 α 的二次函数为

$$\mathbb{E}[(\alpha f_i - X_i)^2] = \alpha^2\mathbb{E}[f_i^2] - 2\alpha\mathbb{E}[f_i X_i] + \mathbb{E}[X_i^2],$$

其最小值在 $\alpha^\star = \mathbb{E}[f_i X_i]/\mathbb{E}[f_i^2]$ 处取得。现在对 $f_i \in \mathbb{R}[Y]_{\leqslant D}$ 取下确界：

$$\mathrm{MMSE}_{X_i}^{\leqslant D} = \inf_{f_i \in \mathbb{R}_{\leqslant D}} \inf_{\alpha \in \mathbb{R}} \mathbb{E}[(\alpha f_i(Y) - X_i)^2] = \mathbb{E}[X_i^2] - \sup_{f \in \mathbb{R}_{\leqslant D}} \frac{\mathbb{E}[f_i(Y)X_i]^2}{\mathbb{E}[f_i(Y)^2]}.$$

最后，通过令 $f_i := f_i/\sqrt{\mathbb{E}[f_i(Y)^2]}$ 对 f_i 进行归一化（当 $\mathbb{E}[f(Y)^2] \neq 0$ 时），于是

$$\sup_{f_i \in \mathbb{R}_{\leqslant D}} \frac{\mathbb{E}[f_i(Y)X_i]^2}{\mathbb{E}[f_i(Y)^2]} = \Big(\sup_{\substack{f_i \in \mathbb{R}_{\leqslant D} \\ \mathbb{E}[f_i(Y)^2]=1}} \mathbb{E}[f_i(Y)X_i] \Big)^2 = \big(\mathrm{Corr}_{P_0,i}^{\leqslant D}\big)^2.$$

代入可得：

$$\inf_{f_i \in \mathbb{R}[Y]_{\leqslant D}} \mathbb{E}\big[(f_i(Y) - X_i)^2\big] = \mathbb{E}[X_i^2] - \big(\mathrm{Corr}_{P_0,i}^{\leqslant D}\big)^2.$$

对 i 求和可得

$$\mathrm{MMSE}_X^{\leqslant D} = \sum_{i=1}^n \mathbb{E}[X_i^2] - \sum_{i=1}^n \big(\mathrm{Corr}_{P_0,i}^{\leqslant D}\big)^2 = \mathbb{E}\|X\|^2 - \sum_{i=1}^n \big(\mathrm{Corr}_{P_0,i}^{\leqslant D}\big)^2.$$

对于第二个结论，考虑任意的 $(f_1, \dots, f_n) \in \mathbb{R}[Y]_{\leqslant D}^n$。根据 $\mathsf{Corr}_{P_0}^{\leqslant D}{}_{,i}$ 的定义以及柯西–施瓦茨不等式，对每个 i，我们有

$$\mathbb{E}[f_i(Y)X_i] \leqslant \sqrt{\mathbb{E}[f_i(Y)^2]}\mathrm{Corr}_{P_0,i}^{\leqslant D}.$$

对 i 求和并再次应用柯西–施瓦茨不等式可得

$$\mathbb{E}\Big[\sum_{i=1}^n f_i(Y)X_i\Big] \leqslant \sum_{i=1}^n \sqrt{\mathbb{E}[f_i(Y)^2]}\mathrm{Corr}_{P_0,i}^{\leqslant D} \leqslant \sqrt{\mathbb{E}\sum_{i=1}^n f_i(Y)^2}\sqrt{\sum_{i=1}^n \big(\mathrm{Corr}_{P_0,i}^{\leqslant D}\big)^2}.$$

在约束条件 $\mathbb{E}\sum_i f_i(Y)^2 = 1$ 下，我们得到上界

$$\mathrm{Corr}_{P_0}^{\leqslant D} \leqslant \sqrt{\sum_{i=1}^n \big(\mathrm{Corr}_{P_0,i}^{\leqslant D}\big)^2}.$$

为了得到相应的下界，固定 $\varepsilon > 0$。对于每个 i，选取 $g_i \in \mathbb{R}[Y]_{\leqslant D}$，使得 $\mathbb{E}[g_i(Y)^2] = 1$ 且 $\mathbb{E}[g_i(Y)X_i] \geqslant \mathsf{Corr}_{P_0}^{\leqslant D}{}_{,i} - \varepsilon$（根据上确界的定义，这是可能的）。定义

$$f_i(Y) := \frac{\mathrm{Corr}_{P_0,i}^{\leqslant D}}{\sqrt{\sum_{j=1}^n (\mathrm{Corr}_{P_0,j}^{\leqslant D})^2}} g_i(Y).$$

于是 $\mathbb{E}\sum_i f_i(Y)^2 = 1$，从而

$$\begin{aligned}
\mathbb{E}\Big[\sum_{i=1}^n f_i(Y)X_i\Big] &= \frac{1}{\sqrt{\sum_j (\mathrm{Corr}_{P_0,j}^{\leqslant D})^2}} \sum_{i=1}^n \mathrm{Corr}_{P_0,i}^{\leqslant D}\, \mathbb{E}[g_i(Y)X_i] \\
&\geqslant \frac{1}{\sqrt{\sum_j (\mathrm{Corr}_{P_0,j}^{\leqslant D})^2}} \sum_{i=1}^n \mathrm{Corr}_{P_0,i}^{\leqslant D}\, (\mathrm{Corr}_{P_0,i}^{\leqslant D} - \varepsilon) \\
&= \sqrt{\sum_{i=1}^n (\mathrm{Corr}_{P_0,i}^{\leqslant D})^2} - \varepsilon \frac{\sum_{i=1}^n \mathrm{Corr}_{P_0,i}^{\leqslant D}}{\sqrt{\sum_{j=1}^n (\mathrm{Corr}_{P_0,j}^{\leqslant D})^2}}.
\end{aligned}$$

对 f_i 取上确界，再令 $\varepsilon \downarrow 0$，得到

$$\mathrm{Corr}_{P_0}^{\leqslant D} \geqslant \sqrt{\sum_{i=1}^n \big(\mathrm{Corr}_{P_0,i}^{\leqslant D}\big)^2}.$$

结合上界即可证明第二个结论。 □

命题的证明 7.3。 证明依赖于概率论家厄米多项式的生成函数刻画。由 [MOS13][第 253], 页，对任意 x, $t \in \mathbf{R}$ 和 $\alpha \in \mathbf{N}$, $\sum_{\alpha=0}^{\infty} \frac{t^{\alpha}}{\alpha!} H_{\alpha}(x) = e^{tx-t^2/2}$, 多元生成函数为：

$$G(t,x) \coloneqq \exp\left(\langle t,x\rangle - \frac{1}{2}|t|^2\right) = \prod_{i=1}^{N} e^{t_i x_i - t_i^2/2} = \prod_{i=1}^{N}\left(\sum_{\alpha_i=0}^{\infty} \frac{t_i^{\alpha_i}}{\alpha_i!} H_{\alpha_i}(x_i)\right) = \sum_{\alpha\in\mathbf{N}^N} \frac{t^{\alpha}}{\alpha!} H_{\alpha}(x), \quad \forall t \in \mathbb{R}^N.$$

我们在平移后的自变量 $x = z+\mu$ 处计算生成函数。根据内积的线性性，$\langle t,\ z+\mu\rangle = \langle t,\ z\rangle + \langle t,\ \mu\rangle$，因此可以将指数项提取出来：

$$G(t, z+\mu) = \exp\left(\langle t,z\rangle + \langle t,\mu\rangle - \frac{1}{2}|t|^2\right) = \exp\left(\langle t,z\rangle - \frac{1}{2}|t|^2\right) \cdot \exp\left(\langle t,\mu\rangle\right). \tag{B.5}$$

现在，我们将右侧的两个因子分别展开为幂级数。第一个因子是 $H_{\gamma}(z)$ 的生成函数，第二个因子是标准指数级数：

$$G(t,z+\mu) = \left(\sum_{\gamma\in\mathbf{N}^N} \frac{t^{\gamma}}{\gamma!} H_{\gamma}(z)\right)\left(\sum_{\beta\in\mathbf{N}^N} \frac{t^{\beta}}{\beta!}\mu^{\beta}\right). \tag{B.6}$$

应用多元幂级数的柯西乘积公式，我们合并这两个求和式。引入多重指标 $\alpha = \gamma + \beta$，这意味着 $\beta = \alpha - \gamma$。条件 $\beta \in \mathbf{N}^N$ 要求$\gamma \leqslant \alpha$ （按分量）。因此，我们将（B.6）改写为：

$$G(t,z+\mu) = \sum_{\alpha\in\mathbf{N}^N} t^{\alpha}\left(\sum_{\gamma\leqslant\alpha} \frac{H_{\gamma}(z)}{\gamma!}\frac{\mu^{\alpha-\gamma}}{(\alpha-\gamma)!}\right). \tag{B.7}$$

另一方面，根据生成函数在（B）中的定义，并将自变量取为 $z+\mu$，左侧为：

$$G(t,z+\mu) = \sum_{\alpha\in\mathbf{N}^N} t^{\alpha}\left(\frac{H_{\alpha}(z+\mu)}{\alpha!}\right). \tag{B.8}$$

由于幂级数表示是唯一的，我们令（ t^{α} ）中的系数与（B.7）和（B.8）中的对应系数相等：

$$\frac{H_{\alpha}(z+\mu)}{\alpha!} = \sum_{\gamma\leqslant\alpha} \frac{H_{\gamma}(z)}{\gamma!(\alpha-\gamma)!}\mu^{\alpha-\gamma}. \tag{B.9}$$

两边同乘 $\alpha!$，得到：

$$H_{\alpha}(z+\mu) = \sum_{\gamma\leqslant\alpha} \frac{\alpha!}{\gamma!(\alpha-\gamma)!} H_{\gamma}(z)\mu^{\alpha-\gamma}. \tag{B.10}$$

注意到多重指标二项式系数 $\binom{\alpha}{\gamma} = \frac{\alpha!}{\gamma!(\alpha-\gamma)!}$，即可得到所述恒等式。 □