arXiv:2602.09959v1 [math.ST] 10 Feb 2026

通过调和分析学习多指标模型中的统计-计算权衡

Hugo Latourelle-Vigeant 和 Theodor Misiakiewicz

耶鲁大学统计与数据科学系

2026年2月11日

摘要

我们研究多指标模型的学习问题，其中标签仅通过未知的 s维投影 $\boldsymbol{W}_*^{\mathsf{T}}\boldsymbol{x} \in \mathbb{R}^s$依赖于输入 $\boldsymbol{x} \in \mathbb{R}^d$。利用该问题在正交群 $\mathcal{O}_d$作用下的等变性，我们得到了球对称输入下多指标模型学习复杂度的精确调和分析刻画，进一步细化并推广了以往针对高斯分布的分析。具体而言，我们在统计查询（SQ）和低次多项式（LDP）框架下推导了统计与计算复杂度下界。这些下界可自然地按球谐子空间分解。受此分解启发，我们构造了一系列基于调和张量展开的谱算法，能够逐步恢复潜在方向，并（近乎）达到这些 SQ 和 LDP 下界。根据调和次数序列的选择，这些估计量可以实现样本复杂度与运行时间复杂度之间广泛的权衡。从技术角度看，我们的结果建立在 $\mathcal{O}_d$ 在 $L^2(\mathbb{S}^{d-1})$ 上作用的半单分解，以及球谐函数与无迹对称张量之间的交织同构之上。

目录

1 引言 **3**

1.1 学习多指标模型 .3 1.2 主要结果概述 .5 1.2.1 学习多指标模型的下界：跃迁复杂度 .6 1.2.2 通过迭代调和张量展开学习多指标模型 .8 1.2.3 示例 .10 1.2.4 论文结构 .11 1.3 其他相关工作 .12 1.4 记号.13

2 技术背景 **13**

2.1 无迹对称张量、函数空间与群作用 .14 2.2 球谐函数与交织算子 .15

3 下界与跃迁复杂度 **18**

3.1 统计查询与低次框架 .18 3.2 弱恢复下界：对齐复杂度 .19 3.3 强恢复下界：跃迁复杂度 .21

4 迭代谐波张量展开算法 **23**
4.1 单步谐波张量展开 .234.1.1 预备知识 .244.1.2 调和张量展开算法 .254.2 多步过程 .29

5 应用 **32**
5.1 学习高斯多指标模型 .325.2 学习方向 MIM.345.3 其他应用 .34

6 讨论 **35**

A 下界的证明 **45**

B 分析 y调和张量展开的分析 g alg算法 **48**
B.1 张量展开的一步：非对称情形 .48B.1.1 引理 1 和引理 2 的证明 .50B.1.2 引理 3 的证明 .
53B.2 张量展开的一步：对称情形 .58B.2.1 引理 7 的证明 .59B.3 多步过程：迭代张量展开 .
61B.4 算法的运行时间 .64

C 应用的详细内容与证明 **65**
C.1 高斯和方向性 MIM 的特化 .65C.1.1 埃尔米特张量及其调和分解 .65C.1.2 高斯 MIM .
68C.1.3 方向 MIM.73C.2 球面上的多项式学习.74

D 补充技术背景 78
D.1 对称无迹张量.78D.2 张量上的正交群作用.81D.3 无迹对称张量与球谐函数之间的同构 . . . .
. . 83D.4 张量积表示 .87D.5 矩阵系数的超收缩性 .89D.6 矩阵集中不等式 .90

1 引言

过去几十年来，统计学和学习理论的一个重点是理解计算瓶颈 在高维学习中的表现——也就是说，某项任务何时能够以计算高效的方式解决，而不仅仅是以统计高效的方式解决。事实上，在许多情形下，计算可处理性的限制远比统计可行性严格：存在广泛的参数区域，在其中学习在信息论意义上是可行的，但尚无已知的多项式时间算法能够成功——这就是所谓的计算—统计差距 [KWB19，BEAH+22，Wei25]。更一般地，我们会观察到计算—统计权衡：额外的计算能力可以弥补样本数量的不足，反之亦然。理解此类权衡何时出现以及如何量化，已成为学习理论界的重要目标。

针对这些问题，主要有两条研究脉络：（1）高维推断，包括植入团 [AKS98,BHK+19]、稀疏主成分分析 [BR13,BB19], 、张量主成分分析 [MR14,HKP+17]以及社区检测 [HWX15, Abb18]。这些研究确定了信噪比阈值，将（据猜想）不存在多项式时间算法能够成功的区域，与存在高效方法的区域区分开来。（2）学习理论，重点关注结构性假设如何实现函数类学习的计算可处理性，例如奇偶函数、稀疏函数、联合函数或决策树 [Man94, Kea98, Fel07, SSBD14, FGR+17, DV21]。尤其是，近期研究 [ABAB+21、AAM22、AAM23、JMS24、BBPV25、TDD+24] 强调了阶梯型结构在函数谱分解（例如傅里叶或埃尔米特分解）中的作用，算法可以利用这种结构，按照难度递增的顺序高效学习各个分量。

本文旨在将这两种方法结合起来。我们聚焦于高维多指标模型这一经典学习问题。近年来，该问题重新受到关注，部分原因在于它与神经网络的联系；例如，参见 [DH18、CM20、AGJ21、CKM22、DLS22、AAM23、GGKS23、TDD+24、DIKZ25、DIKR25、DLB25] 及其中的参考文献。我们刻画了这些模型中的计算—统计权衡，并证明帕累托前沿（可实现的权衡）可能高度不均匀且不连续。我们的主要贡献有两点：

(1) 我们在统计查询（SQ）框架下建立了查询复杂度下界（运行时间的替代指标），并在低次多项式（LDP）框架下建立了样本复杂度下界（低于该最小样本量时，没有多项式时间算法能够成功）；适用于输入分布任意球对称的多指标模型。以往研究主要关注高斯输入，其分析在很大程度上依赖于高斯分布特有的性质。

（2）我们提出了一系列迭代调和张量展开算法，按顺序恢复支撑集，并将近期算法推广自 [JKMS25、DLB25]。

我们证明，这些估计量在局部差分隐私和统计查询框架下，达到了最优样本复杂度和（近乎）最优查询复杂度。此外，通过选择不同的调和次数序列，这些估计量可以实现样本复杂度与运行时间之间的中间权衡。

1.1 多指标模型的学习

多指标模型（MIM）是定义在 $(y,\ \boldsymbol{x}) \in \mathcal{Y} \times \mathbb{R}^d$ 上的一种联合分布，其形式为

$$(y, \boldsymbol{x}) \sim \mathbb{P}_{\rho}^{\boldsymbol{W}_*}: \qquad \boldsymbol{x} \sim \mu, \qquad y|\boldsymbol{x} \sim \rho(\cdot|\boldsymbol{W}_*^{\mathsf{T}}\boldsymbol{x}), \tag{1}$$

其中，$\boldsymbol{W}_* \in \mathbb{R}^{d\times \mathsf{s}}$ 是未知的秩为 s 的支撑集，$\rho(\cdot|\boldsymbol{t}) \in \mathcal{P}(\mathcal{Y})$，$\boldsymbol{t} \in \mathbb{R}^{\mathsf{s}}$是链接函数。因此，响应 y 仅通过协变量 $\boldsymbol{x}$ 在其 s 维投影 $\boldsymbol{W}_*^{\mathsf{T}}\boldsymbol{x}$上的投影来依赖该协变量。

假设隐藏子空间的维数 s 固定，且远小于环境维数 d。当 s $=1$ 时，这种情形也称为单指标模型（SIMs）或广义线性模型。

由于简化性和灵活性，多指标模型数十年来一直在统计学和学习理论中发挥着核心作用 [NW72，McC84，HS89，Li91，DJS08，KS09a，KKSK11，BH25]。利用样本学习（1）的问题历史悠久且内容丰富，关于这一领域的综述，请参阅近期的综述 [BH25] 。下文将简要总结与本文相关的一些重要先验工作（另见第1.3节）。

从较高层面看，学习多指标模型可视为一项分两步完成的任务：（i）恢复子空间 $\boldsymbol{W}_*$（一个高维问题）；（ii）在这个低维子空间上估计链接函数 ρ。在温和的正则条件下，这个问题在信息论意义上很容易：通过在一个 ε-网格上进行穷举搜索，只需 $O_d(d/\varepsilon^2)$ 个样本即可将$\boldsymbol{W}_*$ 恢复到精度 ε [DPVLB24, BH25]。当然，这类方法在高维情况下计算上难以实现。多年来，人们提出了几种高效方法，包括线性 [Bri82]和基于矩的估计量[DC00, LW07, KLV21]，主海森方向 [Li92,LL20, MM18]，以及梯度外积张成空间 [Sam93, HJS01, MZST06, TWKS14]。一个典型例子（例如，参见 [CM20]）通过取顶部特征向量的张成空间来估计支撑集

$$\widehat{\boldsymbol{M}} := \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}(y_i)(\boldsymbol{x}_i\boldsymbol{x}_i^{\mathsf{T}} - \mathbf{I}_d). \tag{2}$$

然而，这种估计器以及其他类似方法，只有在链接函数满足严格条件时才能奏效；一旦这些假设被违背，就可能彻底失效。

近期的一系列研究试图刻画在高斯输入下，使用多项式时间算法学习多指标模型的精确极限 $\mu = \mathcal{N}(0, \mathbf{I}_d)$ [BKM+19，LL20，MM18，DPVLB24，TDD+24，KYM25，DIKZ25，DIKR25，DLB25]。这些研究带来了两个关键洞见：

（a）复杂度由链接函数的埃尔米特展开决定 [DH18，BKM+19，MM18，AGJ21，DPVLB24]：多项式时间算法需要 $\Theta_d(d^{\max(1,k_*/2)})$ 个样本才能成功（在统计查询和局部差分隐私框架下），其中 k_* 表示第一个非零埃尔米特系数的阶数，即所谓的生成（跃迁）指数 [DPVLB24，DLB25，DIKR25，DIKZ25]。

(b) 最优恢复需要多步过程：单步估计量（例如（2））可能被证明并非最优（例如，参见 [DIKZ25]）。一个简单的示例是

$$y = z_1 + z_1z_2z_3, \qquad z_i := \mathrm{sign}(\langle \boldsymbol{w}_{*,i}, \boldsymbol{x}\rangle), \qquad \langle \boldsymbol{w}_{*,i}, \boldsymbol{w}_{*,j}\rangle = \delta_{ij}. \tag{3}$$

一步方法必须利用三次项同时拟合全部三个方向，并且需要 $\Theta_d(d^{3/2})$个样本。相比之下，两步程序先根据线性项估计 z_1，然后根据三次项估计 z_2z_3，总共只需 $\Theta_d(d$个样本。这类程序的复杂度由跃迁复杂度刻画：最优多步恢复过程中最困难阶段的代价 [AAM22、AAM23、BBPV25、JMS24、DLB25、DIKR25]。

本文重新探讨高维多指标模型的学习问题，并考虑一种一般的球不变输入分布[1] μ。在这种设定下，模型（1）关于正交群 $\mathcal{O}_d$是等变的，即（y，$\boldsymbol{x}$）$\sim \mathbb{P}_\rho^{\boldsymbol{W}_*}$意味着（$y$，$g^{-1}\cdot\boldsymbol{x}$）$\sim \mathbb{P}_\rho^{g\cdot\boldsymbol{W}_*}$对所有 $g \in \mathcal{O}_d$都成立，而恢复 $\boldsymbol{W}_*$ 之所以困难，是因为需要打破这种旋转对称性。只有链接函数中在 $\mathcal{O}_d$ 作用下发生非平凡变换的分量，才包含有关潜在子空间的信息。

[1] By 通过Hunt-Stein型论证，这类分布是 *least favorable* 用于等变估计 .

这一视角自然引出对 $\mathcal{O}_d$ 在模型上作用的分析——具体而言，是将其分解为球谐函数，它们是 $\mathcal{O}_d$的不可约表示。直观而言，估计问题的复杂度将由最先出现的对称性破缺谐波分量决定。我们认为，即使在高斯情形下，这一等变视角也为研究学习 MIM 的复杂度提供了一种尤为自然的方法。第二作者近期的一篇论文 [JKMS25] 将这一视角用于单指标模型（$\mathsf{s} = 1$），并表明，相较于厄米特基，采用球谐基进行展开能更有原则地推导出最优算法，同时阐明并揭示了若干新现象（更多讨论见第1.3节）。

我们建立了球对称输入下多指标模型学习复杂度的调和分析刻画。具体而言，我们的工作沿两个方向拓展了近期关于学习高斯多指标模型的研究 [TDD+24，DIKR25、DIKZ25、DLB25]：

任意球不变输入分布。我们的分析利用了 $\mathcal{O}_d$的群论性质，由此自然地导出上界与下界。我们的结果可以直接特化到高斯情形，同时也刻画了高斯情形之外出现的行为（另请参见 [JKMS25]中的讨论）。

区分统计复杂度与运行时间复杂度。以往工作要么关注多项式时间算法的最优样本复杂度（通过局部差分隐私） [DPVLB24、DIKZ25、DLB25] ，要么关注最优查询复杂度（通过统计查询） [AAM23、JMS24、BBPV25], ，因而不同论文对跃迁复杂度的定义各不相同。我们表明，这种差异是内在的：这两种框架捕捉的是根本不同的障碍，并导出两个不同的量——样本跃迁复杂度和查询跃迁复杂度——它们分别描述了在各自资源维度上最优的两种过程。一般而言，两者并不一致；我们的下界表明，在学习多指标模型时，没有算法能够同时达到最优样本复杂度（在多项式时间算法中）和最优运行时间复杂度。

从技术角度来看，我们的结果利用了 $\mathcal{O}_d$ 对 $L^2(\mathbb{S}^{d-1})$ 的作用在球谐函数不可约子空间上的半单分解，以及球谐函数与无迹对称张量之间的交织同构。这种基于对称性的分析自然地拓展至多指标模型和 $\mathcal{O}_d$，适用于紧群作用下更一般的等变学习问题。我们将在后续论文中进一步探讨这一方向 [JKMS26]。

1.2 主要结果概述

令 $\mathbb{S}^{d-1}$ 表示 $\mathbb{R}^d$ 中的单位球面，令 $\tau_d := \mathrm{Unif}(\mathbb{S}^{d-1})$ 表示球面上的均匀测度。对于 $\mathsf{s} \in [d]$，我们将 $\tilde{\tau}_{d,\mathsf{s}}$ 记为 $\boldsymbol{z} \sim \tau_d$ 前 s 个坐标的边缘分布，并将 $\mathrm{St}_{\mathsf{s}}(\mathbb{R}^d) = \{\boldsymbol{M} \in \mathbb{R}^{d\times \mathsf{s}} : \boldsymbol{M}^\mathsf{T}\boldsymbol{M} = \mathbf{I}_\mathsf{s}\}$ 记为 $\mathbb{R}^d$ 中正交归一 s-标架构成的斯蒂费尔流形。

本文始终考虑以下一类球面多指标模型。

定义 1（球面多指标模型）。指标为 $\mathsf{s} \in \mathbb{N}$ 的一个球面对称多指标模型是定义在 $(\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{Y} \times \mathbb{S}^{d-1}$ 上的一个联合分布 $\mathbb{P}^{\boldsymbol{W}_*}_{\nu_d}$，由马尔可夫核 $\nu_d(\cdot|\boldsymbol{t}) \in \mathcal{P}(\mathcal{Y})$、$\boldsymbol{t} \in \mathbb{R}^\mathsf{s}$ 以及一个正交规范s标架 $\boldsymbol{W}_* \in \mathrm{St}_\mathsf{s}(\mathbb{R}^d)$ 指定，使得

$$(\boldsymbol{y}, \boldsymbol{z}) \sim \mathbb{P}^{\boldsymbol{W}_*}_{\nu_d} : \qquad \boldsymbol{z} \sim \tau_d \quad \text{and} \quad \boldsymbol{y}|\boldsymbol{z} \sim \nu_d(\cdot|\boldsymbol{W}_*^\mathsf{T}\boldsymbol{z}).$$

我们允许响应 $\boldsymbol{y}$ 取值于任意可测空间 $\mathcal{Y}$，并用 ν_d^Y 表示 $\mathbb{P}^{\boldsymbol{W}_*}_{\nu_d}$ 下 $\boldsymbol{y}$ 的边缘分布（与 $\boldsymbol{W}_*$ 独立）。注意，对所有 $g \in \mathcal{O}_d$，$(\boldsymbol{y}, g^{-1}\cdot\boldsymbol{z}) \sim \mathbb{P}^{g\cdot\boldsymbol{W}_*}_{\nu_d}$；为简化起见，我们通常省略上标，记作 $\mathbb{P}_{\nu_d}$。

定义 1 中涵盖了具有球面对称输入的经典多指标模型表述（1） :

例 1（球不变输入分布）。设 $\mu \in \mathcal{P}(\mathbb{R}^d)$ 在正交变换下保持不变，即 $g_{\#}\mu = \mu$ 对所有 $g \in \mathcal{O}_d$ 均成立。此类分布具有如下极分解$\boldsymbol{x} = r\boldsymbol{z}$，其中 $r = \|\boldsymbol{x}\|_2 \sim \mu_r$ 与 $\boldsymbol{z} = \boldsymbol{x}/\|\boldsymbol{x}\|_2 \sim \tau_d$相互独立。于是，通过定义（1）中的 *MIM* 可改写为球面多指标模型：$\boldsymbol{y}: = (y, r)$ 以及

$$\boldsymbol{z} \sim \tau_d, \qquad \boldsymbol{y} = (y, r) | \boldsymbol{z} \sim \nu_d(\mathrm{d}\boldsymbol{y} | \boldsymbol{W}_*^{\mathsf{T}} \boldsymbol{z}) := \rho(\mathrm{d}y | r\boldsymbol{W}_*^{\mathsf{T}} \boldsymbol{z}) \mu_r(\mathrm{d}r).$$

高斯多指标模型对应于设定 $\mu_r : = \chi_d$。

给定从球对称多指标模型 $\mathbb{P}_{\nu_d}^{\boldsymbol{W}_*}$ 中抽取的独立同分布样本 $\{(\boldsymbol{y}_i, \boldsymbol{z}_i)\}_{i \le n}$，我们考虑恢复潜在子空间 $\mathsf{span}(\boldsymbol{W}_*)$ 的问题。我们关注高维情形：环境维数 d 很大，而指标 s 保持固定（或随 d 缓慢增长[2]）。我们的目标是刻画算法完成该恢复任务所需的样本量 n 和运行时间 T。具体而言，我们研究：

(i) 多项式时间算法所能达到的最优 样本复杂度（从推测的局部差分隐私下界的意义上而言）。稍微滥用术语，我们将其称为样本最优 复杂度。相比之下，纯粹的信息论 样本最优复杂度——对于不受限制的算法而言——通常为 $\Theta_d(d)$。

(ii) 最优的运行时间复杂度，我们通过统计查询框架中的查询复杂度对其进行启发式刻画。

尽管这些下界（如统计查询/局部差分隐私文献中的惯例）必然是推测性的，但我们提出了一系列迭代算法，实现了相匹配的上界。

备注 1.1 (我们知道 ν_d吗？)。为简化起见，我们假设 ν_d 已知且固定；这对应于完全贝叶斯设定，其中模型固定，且 $\boldsymbol{W}_*$ 从 $\mathsf{Stf}_{\mathsf{s}}(\mathbb{R}^d)$ 中均匀抽取。显然，当 ν_d 未知时，我们的下界仍然成立；而我们的算法依赖于对数据进行一般变换，也可以扩展到这种设定。受篇幅限制，我们在此不再探讨这一方向。恢复 $\boldsymbol{W}_*$后，可以用 ν_d 分段线性函数或其他非参数方法拟合（例如参见 [DIKR25、DIKZ25]）。

1.2.1 学习多指标模型的下界：跃迁复杂度

我们首先在低阶多项式 [KWB19、Wei25] 和统计查询框架 [Kea98、Rey20]（见第3.1 节的简要概述）中，确立使用多项式时间算法学习球面多指标模型时样本复杂度和运行时间复杂度的下界。

调和分解。我们的下界以 $\mathcal{O}_d$的半单分解为表述，其中 $L^2(\mathbb{S}^{d-1})$分解为球谐子空间：

$$L^2(\mathbb{S}^{d-1}) = \bigoplus_{\ell=0}^{\infty} \mathscr{S}_{d,\ell}, \qquad N_{d,\ell} := \dim(\mathscr{S}_{d,\ell}) = \Theta_d(d^\ell),$$

其中，$\mathscr{S}_{d,\ \ell}$ 表示次数为ℓ的球谐函数构成的不可约子空间。对于每个 $\ell \ge 0$，子空间$\mathscr{S}_{d,\ \ell}$ 可通过酉等价识别为 ℓ阶无迹对称张量空间，记作 $\mathsf{TSym}_\ell(\mathbb{R}^d)$。更准确地说，存在一个次数为$\ell$ 的调和张量$\mathcal{H}_{d,\ \ell} : \mathbb{S}^{d-1} \to \mathsf{TSym}_\ell(\mathbb{R}^d)$，其分量均为次数为$\ell$ 的球谐函数，使得映射

$$\Phi_{d,\ell} : \mathrm{TSym}_\ell(\mathbb{R}^d) \to \mathscr{S}_{d,\ell}, \qquad \Phi_{d,\ell}(\boldsymbol{A})(\boldsymbol{z}) = \langle \boldsymbol{A}, \mathcal{H}_{d,\ell}(\boldsymbol{z}) \rangle_{\mathsf{F}}$$

[2] Ou对于固定的 ν_d，这些 r 保证将以非渐近行为成立，其中 ν_d 本身可以依赖于 d .

是一个交织等距同构：对于所有 $\boldsymbol{A},\ \boldsymbol{B} \in \mathrm{TSym}_\ell(\mathbb{R}^d)$、$\boldsymbol{z} \in \mathbb{S}^{d-1}$ 和 $g \in \mathcal{O}_d$，

$$\langle \Phi_{d,\ell}(\boldsymbol{A}), \Phi_{d,\ell}(\boldsymbol{B}) \rangle_{L^2} = \langle \boldsymbol{A}, \boldsymbol{B} \rangle_{\mathsf{F}}, \qquad \Phi_{d,\ell}(\boldsymbol{A})(g^{-1} \cdot \boldsymbol{z}) = \Phi_{d,\ell}(g \cdot \boldsymbol{A})(\boldsymbol{z}).$$

弱恢复的下界。 我们首先考虑对信号子空间 $\boldsymbol{W}_*$进行弱恢复 的任务，即实现优于随机猜测的性能。对于每个 $\ell \geq 1$，定义 ν_d 的第 ℓ个调和系数为

$$\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}) := \mathbb{E}_{\mathbb{P}_{\nu_d}} \left[\mathcal{H}_{d,\ell}(\boldsymbol{z}) \middle| \boldsymbol{y} \right] \in \mathrm{TSym}_\ell(\mathbb{R}^d),$$

并记 $\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2 := \mathbb{E}_{\boldsymbol{y} \sim \nu_d^Y} [\|\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y})\|_{\mathsf{F}}^2]$。我们给出弱恢复所需样本复杂度 n （在局部差分隐私下）和运行时间 $\mathtt{T}$（在统计查询下）的以下下界（定理1）：

$$n \gtrsim \inf_{\ell \geq 1} \frac{d^{\ell/2}}{\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2}, \qquad \mathtt{T} \gtrsim \inf_{\ell \geq 1} \frac{d^{\ell}}{\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2}. \tag{4}$$

这些下界自然地分解到各个不可约子空间上。下确界中的每一项都代表了仅限于使用次数为-ℓ 的球谐函数的算法所面临的下界；这些下界基本上是紧的：我们设计了基于对 $\mathcal{H}_{d,\ell}(\boldsymbol{z})$进行张量展开的谱估计器，对每个 ℓ，这些估计器几乎都能达到相应的下界。注意，$\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2 \lesssim 1$，且会随着 d趋于零，因此这些下界刻画了调和子空间的维数 $N_{d,\ell} = \Theta_d(d^\ell)$与其所携带的关于 $\mathbb{P}_{\nu_d}$的信号强度 $\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2$ 之间的竞争。

由（4），一种自然的弱学习策略是选择使相应下界最小的次数：确定样本最优或运行时间最优的次数

$$\ell_\star^{(s)} \in \arg\min_{\ell \geq 1} \frac{d^{\ell/2}}{\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2}, \qquad \ell_\star^{(q)} \in \arg\min_{\ell \geq 1} \frac{d^{\ell}}{\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2}, \tag{5}$$

并应用相应的张量展开算法。我们始终有 $\ell_\star^{(s)} \geq \ell_\star^{(q)}$。若 $\ell_\star^{(s)} = \ell_\star^{(q)}$，则张量展开估计器可同时达到猜想的样本最优复杂度和（几乎）运行时间最优复杂度。相反，当 $\ell_\star^{(s)} > \ell_\star^{(q)}$ 时，我们的下界表明，不存在单一算法能在样本复杂度和运行时间两方面同时达到最优：必须决定优先考虑哪种资源。最后，也可以选择中间次数 $\ell \notin \{\ell_\star^{(s)}, \ell_\star^{(q)}\}$，以实现样本复杂度与运行时间之间的中间权衡。

强恢复的下界。现在我们转而讨论对 $\boldsymbol{W}_*$的强恢复，即以任意高的精度恢复 $\boldsymbol{W}_*$ 的整个张成空间。如（3）所示，张量展开估计器的一步操作可能只能恢复 $\boldsymbol{W}_*$的一个子空间，因此需要迭代该过程。

假设在某个阶段，我们已经恢复了一个方向的真子集 $\boldsymbol{U} \subsetneq \boldsymbol{W}_*$，其中$\boldsymbol{U} \in \mathrm{St}_{d,r}(\mathbb{R}^d)$。那么，通过对 $\boldsymbol{U}^\mathsf{T}\boldsymbol{z}$ 进行条件化，我们可以将模型 $(\boldsymbol{y},\ \boldsymbol{z}) \sim \mathbb{P}_{\nu_d}$重新参数化为一个约化的 球面多指标模型 $(\boldsymbol{y}_{\boldsymbol{U}},\ \boldsymbol{z}_{\boldsymbol{U}}) \sim \mathbb{P}_{\nu_{d,\boldsymbol{U}}}$，此时维度为 $d-r$，包含 s 个 $-r$指标。具体而言，令 $\boldsymbol{U}_\perp \in \mathbb{R}^{d \times (d-r)}$ 为 $\boldsymbol{U}$ 的一个正交补，并将输入 $\boldsymbol{z} \in \mathbb{S}^{d-1}$ 分解为

$$\boldsymbol{z} = \boldsymbol{U}\boldsymbol{r}_{\boldsymbol{U}} + \sqrt{1 - \|\boldsymbol{r}_{\boldsymbol{U}}\|_2^2}\boldsymbol{U}_\perp \boldsymbol{z}_{\boldsymbol{U}}, \qquad \boldsymbol{r}_{\boldsymbol{U}} := \boldsymbol{U}^\mathsf{T}\boldsymbol{z} \in \mathbb{R}^r, \qquad \boldsymbol{z}_{\boldsymbol{U}} = \frac{\boldsymbol{U}_\perp^\mathsf{T}\boldsymbol{z}}{\|\boldsymbol{U}_\perp^\mathsf{T}\boldsymbol{z}\|_2} \in \mathbb{S}^{d-r-1},$$

从而使得 $\boldsymbol{r}_{\boldsymbol{U}} \sim \tilde{\tau}_{d,r}$与 $\boldsymbol{z}_{\boldsymbol{U}} \sim \tau_{d-r}$ 独立。令 $\boldsymbol{y}_{\boldsymbol{U}} := (\boldsymbol{y}, \boldsymbol{r}_{\boldsymbol{U}})$。于是，在 $(\boldsymbol{y}, \boldsymbol{z}) \sim \mathbb{P}_{\nu_d}^{\boldsymbol{W}_*}$ 下，二元组 $(\boldsymbol{y}_{\boldsymbol{U}}, \boldsymbol{z}_{\boldsymbol{U}})$ 也是环境维数为 $d-r$ 的球面多指标模型，其 $(\mathsf{s}-r)$ 维信号子空间由 $\boldsymbol{U}_\perp^\mathsf{T}\boldsymbol{W}_*$ 张成，链接函数 $\nu_{d,\boldsymbol{U}}$ 由下式给出

$$\boldsymbol{y}_{\boldsymbol{U}} | \boldsymbol{z}_{\boldsymbol{U}} \sim \nu_{d,\boldsymbol{U}}\left(\mathrm{d}\boldsymbol{y}_{\boldsymbol{U}} \middle| \boldsymbol{W}_*^\mathsf{T}\boldsymbol{U}_\perp \boldsymbol{z}_{\boldsymbol{U}}\right) := \nu_d\left(\mathrm{d}\boldsymbol{y} \middle| \boldsymbol{W}_*^\mathsf{T}\boldsymbol{z}\right)\tilde{\tau}_{d,r}(\mathrm{d}\boldsymbol{r}_{\boldsymbol{U}}). \tag{6}$$

因此，我们可以将强恢复归结为这样一列约化球面MIM的弱恢复。对每个 $\ell \geq 1$，定义相应的调和系数

$$\boldsymbol{\xi}_{\boldsymbol{U},\ell}(\boldsymbol{y}_{\boldsymbol{U}}) := \mathbb{E}_{\mathbb{P}_{\nu_d}}\left[\mathcal{H}_{d-r,\ell}(\boldsymbol{z}_{\boldsymbol{U}})\middle|\boldsymbol{y}_{\boldsymbol{U}}\right] \in \mathrm{TSym}_\ell(\mathbb{R}^{d-r}).$$

将弱恢复下界（4）应用于约化模型，可得到依赖于$d-r$ 和 $\boldsymbol{\xi}_{\boldsymbol{U},\,\ell}$ 的下界（对应于作用在输入的子空间 $\boldsymbol{U}_\perp$ 上的子群 $\mathcal{O}_{d-r}$）。

沿任意一系列中间子空间重复这一论证，可得到以下最坏情形复杂度度量，我们称之为样本跃迁和查询跃迁复杂度：[3]

$$\mathsf{sLeap}(\nu_d) = \sup_{\boldsymbol{U}\in\mathrm{Sub}(\boldsymbol{W}_*)} \inf_{\ell\geq 1} \frac{d^{\ell/2}}{\|\boldsymbol{\xi}_{\boldsymbol{U},\ell}\|_{L^2}^2}, \qquad \mathsf{qLeap}(\nu_d) = \sup_{\boldsymbol{U}\in\mathrm{Sub}(\boldsymbol{W}_*)} \inf_{\ell\geq 1} \frac{d^{\ell}}{\|\boldsymbol{\xi}_{\boldsymbol{U},\ell}\|_{L^2}^2}, \tag{7}$$

其中 $\mathrm{Sub}(\boldsymbol{W}_*)$ 表示 $\boldsymbol{W}_*$ 的所有严格子框架组成的集合（此处，$\boldsymbol{W}_*$ 任意固定）。这些度量刻画了为恢复全部 $\boldsymbol{W}_*$而必须解决的最困难的中间子问题。由此得到强恢复下界（定理2）：

$$n \gtrsim \mathsf{sLeap}(\nu_d), \qquad \mathtt{T} \gtrsim \mathsf{qLeap}(\nu_d), \tag{8}$$

分别适用于局部差分隐私（样本）和统计查询（运行时间）框架。

多步恢复算法。跃迁复杂度自然地引出了一类多步算法。固定一列调和次数 $\ell_1,\ \ell_2,\ \ldots \geq 1$。经过 t 步后，假设我们已恢复 $\boldsymbol{U}_{\leq t} \subsetneq \boldsymbol{W}_*$。在第 $t+1$步，对约化 MIM $\nu_{d,\ \boldsymbol{U}_{\leq t}}$ 应用次数为ℓ_{t+1} 的调和张量展开算法，以提取新的方向$\boldsymbol{U}_{t+1} \subseteq \boldsymbol{U}_{\leq t,\ \perp}^{\mathsf{T}} \boldsymbol{W}_*$。然后，令[4]

$$\boldsymbol{U}_{\leq t+1} := \boldsymbol{U}_{\leq t} \oplus \boldsymbol{U}_{t+1},$$

并重复此过程，直到 $\boldsymbol{U}_{\leq T} = \boldsymbol{W}_*$。

为了达到样本最优（分别地，运行时间最优）性能，我们在每一步选择样本最优（分别地，运行时间最优）的调和次数（5）。如果两个次数序列相同，那么同一过程便同时达到样本最优和运行时间最优。然而，通常情况下，调和次数序列可能存在显著差异，其对应的子空间恢复序列也会大不相同$\{\boldsymbol{U}_{\leq t}\}$（示例见下文第1.2.3节）。此时，必须在数据高效和计算高效之间做出选择。最后，选择中间的次数序列，可以在这两种资源之间实现中间权衡。

1.2.2 通过迭代调和张量展开学习多指标模型

对于每个调和次数 $\ell \geq 1$，我们提出一种多项式时间算法，可恢复部分信号方向 $\boldsymbol{W}_*$，并且（几乎）达到（4）中的下界。该算法利用球谐函数与无迹对称张量之间的同构，以及一种将高阶张量映射为矩阵的张量展开操作，其思路源自 [MR14] 关于张量主成分分析的开创性工作。

---

[3]只有当内在维度恰好为 s 时，样本跃迁复杂度和查询跃迁复杂度才有定义。若残差 $\boldsymbol{z}_{\boldsymbol{U}}$ 对于任意真子框架 $\boldsymbol{U} \subsetneq \boldsymbol{W}_*$ 均与 y 独立，则跃迁复杂度变为无穷大，这意味着在恢复 $\boldsymbol{U}$ 后，学习过程便会停滞。[4] 此处及下文中，记号 $\boldsymbol{U}_{\leq t} \oplus \boldsymbol{U}_{t+1}$ 采用广义含义：它表示一个正交归一标架，其张成空间等于 $\mathrm{span}(\boldsymbol{U}_{\leq t}) \oplus \mathrm{span}(\boldsymbol{U}_{\leq t,\ \perp}\boldsymbol{U}_{t+1})$；也就是说，我们通过将 $\boldsymbol{U}_{t+1} \in \mathrm{Stf}(\mathbb{R}^{d-\mathrm{s}_{\leq t}})$ 与 $\mathrm{span}(\boldsymbol{U}_{\leq t})^\perp$ 进行等同，把它视为 $\mathbb{R}^d$ 中的一个标架。

该算法取决于展开形状参数 $(a,\ b)\ \in \mathbb{N}^2$ 的选择。对于 $\ell = 1$，我们取 $(a,\ b) = (1,\ 0)$，而对于 $\ell \geq 2$，我们选择满足 $a + b = \ell$ 的整数 $1 \leq a \leq b$。对于次数为 ℓ 的调和张量 $\mathcal{H}_{d,\ \ell}(\boldsymbol{z}) \in (\mathbb{R}^d)^{\otimes \ell}$，我们将其展开为矩阵，并记作 $\mathrm{Mat}_{a,\ b}(\mathcal{H}_{d,\ \ell}(\boldsymbol{z}))$。给定来自球面多指标模型的 n 个样本 $\{(\boldsymbol{y}_i,\ \boldsymbol{z}_i)\}_{i\in[n]}$，以及一个半正定核 $K\colon \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$，该算法计算经验矩阵的主特征向量

$$\widehat{\boldsymbol{M}} = \frac{1}{n^2}\sum_{i,j=1}^{n}(1 - \delta_{a\neq b}\delta_{i=j})K(\boldsymbol{y}_i, \boldsymbol{y}_j)\mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_i))\mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_j))^{\mathsf{T}} \in \mathbb{R}^{d^a\times d^a}, \tag{9}$$

所得特征向量（位于 $\mathbb{R}^{d^a}$ 中）随后与自身收缩，形成 $d \times d$ 矩阵，该矩阵的顶部特征向量定义了估计 $\widehat{\boldsymbol{U}}_0 \in \mathrm{Stf}_{\mathsf{s}_0}(\mathbb{R}^d)$。该过程的完整描述见算法1 （第4.1节）。

虽然该方法的样本复杂度对于 $a \leq b$ 与展开形状无关，但运行时间取决于这一选择：为使计算成本最小化，展开应尽可能接近方形。当 $a = b = \ell/2$ ℓ 为偶数时我们取之，而当 a、b）$=$（（$\ell - 1$）$/2$、（$\ell + 1$）$/2$）在$\ell \geq 3$ 为奇数时取之。在此选择下，对于核 $\widehat{\ }$ 的适当选取，我们得到如下保证（定理 3 与命题 1）。对于 $\ell \geq 2$，估计子空间 $\boldsymbol{U}_0$ 是 $\boldsymbol{W}_*$ 的某个子空间的良好近似，只要

$$n \asymp \frac{d^{\ell/2}}{\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2}, \qquad \mathsf{T} \asymp \frac{d^{\ell+\frac{1}{2}\delta_\ell}}{\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2}\log(d), \tag{10}$$

其中 $\delta_\ell = 0$ 当 ℓ 且 $\delta_\ell = 1$当 ℓ。因此，当 ℓ为偶数时，该算法达到猜想的 LDP 和 SQ 下界（运行时间相差一个对数因子）。当 ℓ为奇数时，运行时间会额外增加 $\sqrt{d}$倍，我们尚未解决能否去掉这一因子的问题。在单指标模型的特殊情形下（$\mathsf{s} = 1$），[JKMS25] 证明，简单的在线 SGD 算法对所有 $\ell \geq 3$ 都能达到最优运行时间 $\mathsf{T} \asymp d^{\ell}/\|\boldsymbol{\xi}_{\emptyset,\ \ell}\|_{L^2}^2$（不含对数因子），但样本复杂度会显著更差 $n \asymp d^{\ell-1}/\|\boldsymbol{\xi}_{\emptyset,\ \ell}\|_{L^2}^2$。

情形 $\ell = 1$ 则更为微妙。此处，下界（4）适用于检测，但对于恢复而言并不总是紧的：在这种情况下可能出现检测-恢复差距。我们的张量展开估计器能够恢复一部分信号方向，其

$$n \asymp \frac{d}{\|\boldsymbol{\xi}_{\emptyset,1}\|_{L^2}^2}, \qquad \mathsf{T} \asymp \frac{d^2}{\|\boldsymbol{\xi}_{\emptyset,1}\|_{L^2}^2}\log(d). \tag{11}$$

当 $\|\boldsymbol{\xi}_{\emptyset,\ 1}\|_{L^2}^2 \asymp 1$时，样本复杂度是紧的，并且达到了恢复的信息论下界。当 $\|\boldsymbol{\xi}_{\emptyset,\ 1}\|_{L^2}^2 \ll 1$时，在对 ν_d作出额外假设的条件下，可以使用改进算法进一步改善上述界；参见第 4.1.2 节。

为了恢复整个支撑集 $\boldsymbol{W}_*$，我们将这一方法扩展为一个多步过程。固定次数序列 $\{\ell_t\}_{t\geq 1}$后，我们对通过条件化于先前已恢复子空间而得到的约化MIM，迭代应用调和张量展开 $\widehat{\boldsymbol{U}}_{\leq t}^{\mathsf{T}}\boldsymbol{z}$。在对 MIM 的条件分布作出稳定性假设后，我们证明每一步的复杂度都与（10）和（11）中的相应单步保证相匹配。因此，选择样本复杂度最优或运行时间最优的次数序列（即在每一步选择最优次数（5））时，该多步过程能够恢复整个支撑集，其复杂度与跃迁复杂度下界（8）相匹配，样本复杂度至多相差一个 $O(\sqrt{d})$ 因子（当最困难的步骤出现在 $\ell = 1$ 时），运行时间至多相差一个 $\widetilde{O}(\sqrt{d})$ 因子（当最困难的步骤出现在奇数 ℓ 时）。

1.2.3 示例

接下来，我们给出几个示例，说明迭代调和张量展开如何学习多指标模型。更多讨论请参见第 5 节。

例 1：高斯单指标模型。这个例子已在 [JKMS25],中得到广泛研究，我们简要回顾其一些性质。考虑一个高斯单指标模型（$\mathsf{s}=1$），其中

$$\boldsymbol{x}\sim\mathcal{N}(0,\mathbf{I}_d),\qquad y|\boldsymbol{x}\sim\rho(\cdot|\langle\boldsymbol{w}_*,\boldsymbol{x}\rangle),\tag{12}$$

其中链接函数 ρ的生成指数为 $k_*\geq 1$ [DPVLB24] （见第5.1节）。根据例1，该模型可改写为球面单指标模型，因此我们可以应用本文的结果。

对于每个满足 $\ell\leq k_*$ 且 ℓ 和 k_* 奇偶性相同的次数，调和系数满足$\|\boldsymbol{\xi}_{\emptyset,\ \ell}\|_{L^2}^2\asymp d^{-(k_*-\ell}{}_)/2$。将这些估计代入（4），可得样本复杂度下界为 $n\asymp d^{k_*/2}$，且在每个满足 $\ell\leq k_*$ 的次数 $\ell\equiv k_*[2]$ 处都能达到该下界。相比之下，运行时间在次数 $d^{(k_*+\ell)/2}$ 处按 ℓ增长。因此，要同时实现样本复杂度最优和运行时间最优，应选择 $\ell\in\{1,2\}$，使其与 k_*具有相同奇偶性。该次数下的调和张量展开（针对 $\ell=1$作了修改；见 [JKMS25]）可达到

$$n\asymp d^{k_*/2},\qquad \mathtt{T}\asymp d^{k_*/2+1}\log d,$$

二者均为最优。在第 5.1 节中，我们将这一分析扩展到含有 $\mathsf{s}>1$ 的高斯多指标模型；一般而言，不存在能够同时达到样本最优和运行时间最优的单一算法。

文献 12 中的式子有一种有趣的变体， [JKMS25] 讨论了这一变体：仅观测归一化后的输入 $\boldsymbol{z}:=\boldsymbol{x}/\|\boldsymbol{x}\|_2$。这在统计学和机器学习中是一种常见的预处理步骤。由此得到的模型不再是高斯单指标模型，但仍是球面单指标模型，因此我们的理论依然适用。在这种情况下，调和系数满足 $\|\boldsymbol{\xi}_{\emptyset,\ \ell}\|_{L^2}^2\asymp d^{-(k_*-\ell}{}_)$。此时，对于所有 d^{k_*}奇偶性相同的 $\ell\leq k_*$，运行时间为 $d^{k_*-\ell/2}$，而样本复杂度的增长阶为 $\ell=k_*$。因此，样本最优和运行时间最优的选择均为 k_*，并且在 $n\asymp d^{k_*/2}$ 处进行张量展开可达到最优的 $\mathtt{T}\asymp d^{k_*+\frac{1}{2}\delta_{k_*}}\log d$。特别地，归一化输入不会改变样本复杂度，但运行时间会变差一个二次因子。这一现象对基于梯度的算法有重要影响；参见 [JKMS25]。

示例 2：奇偶函数。设 $\boldsymbol{z}\sim\tau_d$，并令 $\boldsymbol{W}_*=[\boldsymbol{w}_{*,1},\dots,\boldsymbol{w}_{*,\mathsf{s}}]\in\mathrm{St}_{\mathsf{s}}(\mathbb{R}^d)$ 表示未知的信号子空间。记 $z_i:=\langle\boldsymbol{w}_{*,\ i},\ \boldsymbol{z}\rangle, i\in[\mathsf{s}]$ 为对应于这些信号方向的投影。考虑含有 $\mathsf{s}\geq 2$ 的含噪奇偶模型，其定义如下：

$$y=\mathrm{sign}(z_1z_2\cdots z_{\mathsf{s}})+\varepsilon,\tag{13}$$

其中 $\varepsilon\sim\mathcal{N}(0,\ \omega^2)$ 为独立加性噪声[5]。这是一个球面多指标模型，可以验证[6] 对于所有$\|\boldsymbol{\xi}_{\emptyset,\ \ell}\|_{L^2}^2=0_{\mathsf{s}}$，均有 $\ell<$ ，且$\|\boldsymbol{\xi}_{\emptyset,\ \mathsf{s}}\|_{L^2}^2\asymp 1$。因此，样本最优次数和运行时间最优次数均为 $\ell_\star^{(s)}=\ell_\star^{(q)}=\mathsf{s}$。在次数为 $\ell=\mathsf{s}$ 时应用调和张量展开，一步即可恢复整个支撑集 $\boldsymbol{W}_*$，其

$$n\asymp d^{\mathsf{s}/2},\qquad \mathtt{T}\asymp d^{\mathsf{s}+\frac{1}{2}\delta_{\mathsf{s}}}\log(d).$$

这在局部差分隐私和统计查询框架下，同时实现了最优样本复杂度和近乎最优的运行时间复杂度。

[5]选择这一加性噪声模型是为了简化；对于任何形式为 $y|\boldsymbol{z}\sim\rho($ $y|$ $z_1z_2\cdots z_{\mathsf{s}}$. sign( ) 的链接函数，相同的结论也成立。例如，可以将 [DLB25 命题 4] 的证明从高斯数据推广到球面数据。

示例 3：奇偶函数混合。令 $p \in (0, 1/2)$，并令 $k_0 < k_1 < k_2$ 为整数；为简化起见，我们取偶数。考虑以下响应模型：

$$y = \eta \cdot \mathrm{sign}(z_1 z_2 \cdots z_{k_1}) + (1-\eta) \cdot \mathrm{sign}(z_{k_1-k_0+1} z_{k_1-a+2} \cdots z_{k_2+k_1-k_0}) + \varepsilon,$$

其中 $\eta \sim$服从伯努利分布(p)，且 $\varepsilon \sim \mathcal{N}(0, \omega^2)$。等价地， y是两个奇偶函数的带噪混合：一个大小为 k_1 （概率为 p），另一个大小为 k_2 （概率为 $1-p$），二者共享 k_0 个信号方向。

可以验证，对于 $\ell < k_1$， $\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2 \asymp p^2$ 当 $k_1 \le \ell < k_2$时成立，而当 $\ell \ge k_2$ (ℓ 为偶数)时，$\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2 \asymp 1$ 成立。因此，最优迭代过程对应于以下两种次数序列之一：$(\ell_1,\ \ell_2)$、$(k_1,\ k_2-k_0)$ 或 $(k_2,\ k_1-k_0)$：

- **次数序列$(k_1,\ k_2-k_0)$：** 首先在次数 k_1 处进行张量展开，以恢复对应于较小奇偶函数的 k_1 个方向。以这些方向为条件后，约化模型变为定义在剩余 k_2-k_0 个方向上的奇偶函数，随后可通过在次数 k_2-k_0处进行张量展开来恢复。上述两个步骤的总复杂度为

$$n \asymp d^{k_1/2}/p^2 + d^{(k_2-k_0)/2}, \qquad \mathtt{T} \asymp (d^{k_1}/p^2 + d^{k_2-k_0})\log(d).$$

- **次数序列 $(k_2,\ k_1-k_0)$：** 我们首先通过在次数 k_2处进行张量展开，恢复与较大奇偶性对应的 k_2 个方向。在以这些方向为条件后，我们通过次数 k_1-k_0恢复其余的 k_1-k_0 个方向。由此得到

$$n \asymp d^{k_2/2} + d^{(k_1-k_0)/2}/p^2, \qquad \mathtt{T} \asymp (d^{k_2} + d^{k_1-k_0}/p^2)\log(d).$$

作为一个具体示例，固定 $q \in \mathbb{N}$ 并设 $k_0 = 2q$、$k_1 = 4q$、$k_2 = 8q$ 和 $p = d^{-3q/2}$。那么这两种策略分别得到

$$\begin{aligned} &\textbf{Sample-optimal algorithm at } (k_2, k_1-k_0)\textbf{:} \qquad n \asymp d^{4q}, \qquad \mathtt{T} \asymp d^{8q}\log(d), \\ &\textbf{Runtime-optimal algorithm at } (k_1, k_2-k_0)\textbf{:} \qquad n \asymp d^{5q}, \qquad \mathtt{T} \asymp d^{7q}\log(d). \end{aligned}$$

这些复杂度分别与（7）中的样本跃迁下界和查询跃迁下界相匹配。特别地，运行时间最优策略将运行时间降低了 d^q 倍，代价是样本量增加d^q 倍。我们的下界表明，在该模型中，没有任何单一算法能够同时实现样本复杂度和运行时间的最优缩放。

1.2.4 论文结构

本文其余部分安排如下。在第 2 节中，我们介绍全文将用到的技术背景，包括张量代数、球谐函数、相关群作用以及联系这些对象的交织算子等标准工具。在第 3 节中，我们简要回顾统计查询和局部差分隐私框架，随后为球面 MIM 建立弱恢复（对齐复杂度）和强恢复（跃迁复杂度）的下界。第 4 节介绍一系列迭代调和张量展开算法。我们首先描述单步展开过程，并给出其样本复杂度和运行时间复杂度保证。随后，我们通过迭代单步过程将这一方法扩展为多步算法，并建立球面 MIM 强恢复的学习保证。第 5节将我们的通用结果应用于更具体的模型类别，包括高斯模型和方向多指标模型。最后，我们在第 6节总结全文，并讨论未来的研究方向。

1.3 其他相关工作

下面我们将进一步讨论多指标模型的相关工作，不再重复引言中已经引用的文献。许多经典问题，例如半空间学习 [Vem97、Bau90、KS09b、KSV24、Vem10], 一比特压缩感知 [PV13、ALPV14、GNJN13], 相位恢复 [CLS15、CSV13、NJS13、MM18], 以及固定宽度神经网络学习 [GLM18、BJW19、CDG+23]，都可以视为链接函数取特定形式时的 MIM 特例。这些联系催生了广泛而多样的研究文献，本文不打算在此进行全面综述。

对于单指标情形，早期研究表明，具有单调链接函数的单指标模型能够以线性样本复杂度实现高效学习 [KS09a,KKSK11]。然而，即便是带噪相位恢复 [BKM+19,MLKZ20, MM18]这类简单的非单调示例，也会出现统计-计算差距。对于一般的高斯单指标模型， [DPVLB24] 给出了统计复杂度和计算复杂度的精确刻画（以 d 中的多项式指数表示），并采用生成指数进行衡量；其最优算法基于对埃尔米特张量取偏迹。这一精确刻画随后被推广到球形单指标模型（定义1中的情形 $s = 1$ ） [JKMS25], ，该工作推导出了统计查询和低次下界（类似于定理1），并提出了调和张量展开[7] 和在线随机梯度下降算法，分别实现最优样本复杂度和运行时间。 [JKMS25] 进一步阐明了这种调和分析视角的优势，解释了若干现象：为何损失景观平滑能够达到最优复杂度，为何随机梯度下降并非最优，为何利用输入范数是实现最优运行时间复杂度的必要条件，以及为何在高斯情形之外还会出现额外的统计-计算权衡。

学习多指标模型的图景丰富得多，需要采用自适应的多阶段过程 [AAM22, AAM23, BBPV25, DLB25, DIKR25, DIKZ25]。在此，学习高斯多指标模型的复杂度由（生成式）跃迁复杂度刻画 [AAM23, BBPV25, JMS24, DLB25, DIKZ25]。在（最优的）比例缩放区域内，研究者通过谱方法和近似消息传递（AMP）研究了跃迁指数为二的高斯多指标模型的弱恢复 $n \asymp d$ TDD+24, DDM+25,KYM25]。在比例区域之外， [DLB25] 通过生成式跃迁指数刻画样本复杂度，并提出了一种展开形状为 $(a, b) = (1, \ell - 1)$ 的埃尔米特张量展开估计量，该估计量通过迭代地对已恢复子空间进行条件化，达到低次数下界。采用这一配置时，我们的估计量能够恢复这些保证。然而，至关重要的是，我们证明，采用更“方形”的展开形状 $(a, b) = (\lfloor \ell/2 \rfloor, \lceil \ell/2 \rceil)$ 并选择不同的次数序列，可能会改善运行时间复杂度。这一改动在保持高斯多指标模型最优样本复杂度的同时，改善了运行时间（并且几乎达到统计查询下界）；不过，我们的下界表明，在更广泛的区域中，样本与运行时间的权衡不可避免。更广泛地说，我们的结果将 [DLB25] 的分析扩展到球面多指标模型，给出了明确的查询复杂度（运行时间）下界，并刻画了该工作未涉及的样本与运行时间权衡。在另一项同期工作中， [DIKR25, DIKZ25]为高斯多指标模型开发了一种基于迭代子空间条件化的相关学习过程。尽管这些工作没有刻画 d 中精确的多项式依赖关系，但它们研究了多指标模型的无先验学习，并跟踪了对其他问题参数的依赖性。我们将对无先验设定的分析扩展以及获得同样精细的界限留待未来工作。

除上述研究外，人们还研究了其他估计算法，包括半参数最大似然法 [Hor09, DH18], 切片逆回归 [Li91, BB18], 以及基于梯度的方法 [AGJ21, ASKL23, ZPVB23, ADK+24a, DNGL24, DTA+24, LOSW24, ADK+24b, MW26]；综述请参见 [CM20, DH18, BH25] 。近期研究还探讨了鲁棒学习 [DKTZ22, LKDD24,ZWDD24, WZDD24] 以及互补的困难性结果，包括无先验学习的密码学下界 [DKR23] 和分布无关 PAC 困难性 [DKMR22]。

[7] 对于秩一核 ，估计器（9）在 [JKMS25] 中恢复了球形 SIM 的张量展开估计器 .

1.4 记号

我们用粗体字母（例如， $\boldsymbol{x}$, $\boldsymbol{A}$）表示向量、矩阵和高阶张量，用常规字母（例如， x, A）表示标量。对于正整数 d，记作 $[d]=\{1,2,\dots,d\}$。 $\mathbb{R}^d$ 中的单位球面记为 $\mathbb{S}^{d-1}$。对于有限集 S， $|S|$ 表示其基数。我们用$\langle\cdot,\ \cdot\rangle$ 表示 $\mathbb{R}^d$上的欧几里得内积和范数，用 $\langle\cdot,\ \cdot\rangle_{\mathsf{F}}$ 和 $\|\cdot\|_{\mathsf{F}}$ 表示高阶张量的弗罗贝尼乌斯内积和范数（$\mathbb{R}^d$）$^{\otimes\ell}$。对于可测空间 $\mathcal{Y}$， $\mathcal{P}(\mathcal{Y})$表示 $\mathcal{Y}$上的概率测度集合。对于可测函数 f， $\|f\|_{L^p(\mu)}$ 表示相对于测度 μ的L^p 范数。当 μ 可由上下文确定时，我们只写作 $\|f\|_{L^p}$。特别地，对于张量值随机变量 $\boldsymbol{\xi}$，记作 $\|\boldsymbol{\xi}\|_{L^2}=\|\|\boldsymbol{\xi}\|_{\mathsf{F}}\|_{L^2}$。我们还用 $\|\cdot\|_{\mathrm{op}}$ 表示线性算子的标准算子范数。

对于矩阵 $\boldsymbol{A}$，$\mathsf{span}(\boldsymbol{A})$ 表示由其列张成的线性子空间（更一般地，$\mathsf{span}(\{\boldsymbol{v}_i\})$ 表示一组向量的线性张成）。对于 $1\le s\le d$，$\mathsf{Stf}_s(\mathbb{R}^d)$ 表示 $\mathbb{R}^d$ 中正交 s-标架构成的斯蒂费尔流形。对于 s、$k\in[d]$，其中 $s+k\le d$、$\boldsymbol{U}\in\mathsf{Stf}_s(\mathbb{R}^d)$ 且 $\boldsymbol{V}\in\mathsf{Stf}_k(\mathbb{R}^d)$，当 $\mathsf{span}(\boldsymbol{U})\cap\mathsf{span}(\boldsymbol{V})=\{0\}$ 时，我们将拼接 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 的列所得到的正交 $(s+k)$-标架记为 $\boldsymbol{U}\oplus\boldsymbol{V}$。此外，如果 $\mathsf{span}(\boldsymbol{U})\subseteq\mathsf{span}(\boldsymbol{V})$，则称 $\boldsymbol{U}$ 是 $\boldsymbol{V}$ 的子标架，记作 $\boldsymbol{U}\subseteq\boldsymbol{V}$；当包含关系为真包含时，称其为严格子标架。给定两个正交 s- 标架 $\boldsymbol{U}$、$\boldsymbol{U}'\in\mathsf{Stf}_s(\mathbb{R}^d)$，我们用 $\mathsf{dist}(\boldsymbol{U},\boldsymbol{U}')=\|\boldsymbol{U}\boldsymbol{U}\mathsf{T}-\boldsymbol{U}'(\boldsymbol{U}')^{\mathsf{T}}\|_{\mathrm{op}}$ 表示它们的列空间上正交投影之间的算子范数距离。

除非另有说明， c',c,C',C表示与维数参数（n,d）无关的正数常数，但可能依赖于固定的模型参数（例如 s ℓ以及假设中的常数）。通常， c,c'表示充分小的常数，而 C,C'表示充分大的常数。它们的取值可能逐行变化。对于函数 $f,g:\mathbb{N}\to\mathbb{R}_{\ge0}$，若存在常数 $C>0$，使得对所有充分大的 d，$f(d)=O(g(d))$（分别地， $f(d)=\Omega(g(d))$），则记作$f(d)=O(g(d))$（分别地， $f(d)=\Omega(g(d))$）。若 $f(d)=O(g(d))$ 且 $g(d)=O(f(d))$，则记作 $f(d)=\Theta(g(d))$。记号 $f(d)=\widetilde{O}(g(d))$ 表示，对某个常数 $C>0$，当 d 充分大时， $f(d)\le Cg(d)\log^C(d)$。除非另有说明，常数 C 可能依赖于固定的模型参数和假设中的常数，但不依赖于维数参数（n,d）。我们用 $f(d)\lesssim g(d)$、 $f(d)\gtrsim g(d)$ 和 $f(d)\asymp g(d)$ 分别简写 $f(d)=O(g(d))$、$f(d)=\Omega(g(d))$ 和 $f(d)=\Theta(g(d))$。 $f(d)=o(g(d))$（或 $f(d)\ll g(d)$）表示 $\lim_{d\to\infty}f(d)/g(d)=0$，而 $f(d)=\omega(g(d))$（或 $f(d)\gg g(g)$）表示 $\lim_{d\to\infty}f(d)/g(d)=\infty$。当 f,g 是多变量函数时，我们有时写作 $f=O_d(g)$，表示渐近记号仅针对变量 d，其他变量保持固定；对于 Ω Θ,O,ω,o，也采用类似写法。

2 技术背景

本节回顾无迹对称张量、球谐函数以及它们与正交群表示之间关系的技术背景 $\mathcal{O}_d$。这些性质大多是经典结果，不过其中少数不太常见，因此也在此列出以求完整。关于球谐函数的全面论述，请参阅 [Sze39、Chi11], ；关于$\mathcal{O}_d$ 的表示论及其与无迹对称张量空间的关系，请参阅 [GW00、FH13]。详细陈述、证明及其他性质留待附录D中介绍。

2.1 无迹对称张量、函数空间与群作用

无迹对称张量。令 $(\mathbb{R}^d)^{\otimes \ell}$ 为 $\mathbb{R}^d$ 上的 ℓ 阶张量空间，并赋予其 Frobenius 内积

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle_{\mathsf{F}} = \sum_{i_1,\dots,i_\ell=1}^{d} A_{i_1,\dots,i_\ell} B_{i_1,\dots,i_\ell}, \qquad \boldsymbol{A}, \boldsymbol{B} \in (\mathbb{R}^d)^{\otimes \ell}.$$

对于整数 $p, q, r \geq 0$，满足 $r \leq \min(p, q)$，张量缩并 $\otimes_r \colon (\mathbb{R}^d)^{\otimes p} \times (\mathbb{R}^d)^{\otimes q} \to (\mathbb{R}^d)^{\otimes p+q-2r}$ 定义为

$$(\boldsymbol{A} \otimes_r \boldsymbol{B})_{i_1,\dots,i_{p-r},j_1,\dots,j_{q-r}} = \sum_{s_1,\dots,s_r=1}^{d} A_{i_1,\dots,i_{p-r},s_1,\dots,s_r} B_{s_1,\dots,s_r,j_1,\dots,j_{q-r}}.$$

特别地，$\otimes := \otimes_0$ 对应于标准张量积。对于 $\boldsymbol{M} \in \mathbb{R}^{m \times d}$ 和 $\boldsymbol{A} \in (\mathbb{R}^d)^{\otimes \ell}$，我们将 $\boldsymbol{M}^{\otimes \ell} \boldsymbol{A} \in (\mathbb{R}^m)^{\otimes \ell}$ 记作通过对每个指标应用 $\boldsymbol{M}$ 得到的张量：

$$\left(\boldsymbol{M}^{\otimes \ell} \boldsymbol{A}\right)_{i_1,\dots,i_\ell} = \sum_{j_1,\dots,j_\ell=1}^{d} M_{i_1,j_1} \cdots M_{i_\ell,j_\ell} A_{j_1,\dots,j_\ell}.$$

我们用 $\mathrm{Sym}^\ell(\mathbb{R}^d) \subseteq (\mathbb{R}^d)^{\otimes \ell}$ 表示对称张量子空间，即满足以下条件的张量 $\boldsymbol{A}$：对任意置换 $\sigma \in \mathfrak{S}_\ell$, $A_{i_1,\dots,i_\ell} = A_{i_{\sigma(1)},\dots,i_{\sigma(\ell)}}$。我们将到此子空间的正交投影定义为

$$\mathsf{P}_{\mathsf{sym}}(\boldsymbol{A}) = \frac{1}{\ell!} \sum_{\sigma \in \mathfrak{S}_\ell} \boldsymbol{A}^\sigma, \tag{14}$$

其中 $\boldsymbol{A}^\sigma$ 是由 $A^\sigma_{i_1,\dots,i_\ell} = A_{i_{\sigma(1)},\dots,i_{\sigma(\ell)}}$定义的张量。

对于 $\ell \geq 2$，我们将偏迹算子 $\tau\colon \mathrm{Sym}^\ell(\mathbb{R}^d) \to \mathrm{Sym}^{\ell-2}(\mathbb{R}^d)$ 定义为

$$\tau(\boldsymbol{A})_{i_1,\dots,i_{\ell-2}} = \sum_{j=1}^{d} A_{j,j,i_1,\dots,i_{\ell-2}}. \tag{15}$$

对于对称张量，该定义与所求迹指标的选取无关。于是，无迹对称张量的子空间为

$$\mathrm{TSym}_\ell(\mathbb{R}^d) = \{\boldsymbol{A} \in \mathrm{Sym}_\ell(\mathbb{R}^d) : \tau(\boldsymbol{A}) = 0\}.$$

对于 $\ell = 0$，1，我们令 $\mathrm{TSym}^0(\mathbb{R}^d) = \mathrm{Sym}^0(\mathbb{R}^d) = \mathbb{R}$ 且 $\mathrm{TSym}^1(\mathbb{R}^d) = \mathrm{Sym}^1(\mathbb{R}^d) = \mathbb{R}^d$。到 $\mathrm{TSym}^\ell(\mathbb{R}^d)$ 上的正交投影记为 $\mathsf{P}_{\mathsf{tf}}\colon \mathrm{Sym}^\ell(\mathbb{R}^d) \to \mathrm{TSym}^\ell(\mathbb{R}^d)$；其显式表达式见附录 D.1。最后，用 $\mathsf{L}_{d,\ell}$ 表示 $\mathrm{TSym}^\ell(\mathbb{R}^d)$ 上的线性算子空间，它可与 $\mathrm{TSym}^\ell(\mathbb{R}^d) \otimes \mathrm{TSym}^\ell(\mathbb{R}^d)$ 等同。$\mathsf{L}_{d,\ell}$ 上的算子范数、迹和弗罗贝尼乌斯范数按通常方式定义。

球面上的函数空间。令 $\mathbb{S}^{d-1} = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 = 1\}$ 表示单位球面，τ_d 表示 $\mathbb{S}^{d-1}$ 上的均匀概率测度。我们将 $L^2(\mathbb{S}^{d-1}) := L^2(\mathbb{S}^{d-1}, \tau_d)$ 记为球面上平方可积函数的空间，并赋予其内积和范数

$$\langle f, g \rangle_{L^2(\tau_d)} = \int_{\mathbb{S}^{d-1}} f(\boldsymbol{x}) g(\boldsymbol{x}) \tau_d(\mathrm{d}\boldsymbol{x}), \qquad \|f\|^2_{L^2(\tau_d)} = \langle f, f \rangle_{L^2(\tau_d)}. \tag{16}$$

我们还将考虑乘积空间

$$L^2(\mathcal{V}\times\mathbb{S}^{d-1},\nu\otimes\tau_d)=L^2(\mathcal{V},\nu)\otimes L^2(\mathbb{S}^{d-1},\tau_d),$$

其对应的内积记为 $\langle\cdot,\cdot\rangle_{L^2(}$ $_{L^2(\mathcal{V},\nu)}$ $_{\nu\otimes\tau_d}$ $=)$。为简化起见，当测度可由上下文明确时，我们常写作 $L^2(\mathcal{V})$ (,) 和 $\langle\cdot,\cdot\rangle_{L^2}$。

正交群。令 $\mathcal{O}_d$ 表示 $\mathbb{R}^d$ 上的正交群，将其视为满足 $\boldsymbol{Q}\boldsymbol{Q}^\mathsf{T}=\mathbf{I}_d$ 的正交矩阵 $\boldsymbol{Q}\in\mathbb{R}^{d\times d}$ 的集合。我们用 π_d 表示 $\mathcal{O}_d$ 上的 Haar 概率测度。群 $\mathcal{O}_d$ 自然地作用于 $\mathbb{R}^d$，并通过 $(\mathbb{R}^d)^{\otimes\ell}$ 作用于

$$g\cdot\boldsymbol{x}=\boldsymbol{Q}\boldsymbol{x},\qquad g\cdot\boldsymbol{A}=\boldsymbol{Q}^{\otimes\ell}\boldsymbol{A}.$$

对任意 $0\le \mathsf{s}\le d$，令 $\mathrm{Stf}_\mathsf{s}(\mathbb{R}^d)=\{\boldsymbol{W}\in\mathbb{R}^{d\times\mathsf{s}}:\boldsymbol{W}^\mathsf{T}\boldsymbol{W}=\mathbf{I}_\mathsf{s}\}$ 表示 $\mathbb{R}^d$ 中正交归一 s-标架构成的斯蒂费尔流形，并赋予其均匀测度 $\pi_{d,\mathsf{s}}$。对于 $\boldsymbol{W}\in\mathrm{Stf}_\mathsf{s}(\mathbb{R}^d)$，$\boldsymbol{W}$ 在 $\mathcal{O}_d$ 的左作用下的稳定子子群为

$$\mathcal{O}_d^{\boldsymbol{W}}=\{g\in\mathcal{O}_d:g\cdot\boldsymbol{W}=\boldsymbol{W}\}.\tag{17}$$

该 群 $\mathcal{O}_d^{\boldsymbol{W}}$作为自然的 $\mathcal{O}_d$ 作用的限制作用于 $\mathbb{R}^d$，并作用于张量 .

最后，我们考虑作用于 $L^2(\mathbb{S}^{d-1})$ 和 $L^2(\mathcal{V}\times\mathbb{S}^{d-1})$ 的 $\mathcal{O}_d$ 的酉表示，其作用方式如下：

$$[\rho(g)\cdot f](\boldsymbol{z})=f(g^{-1}\cdot\boldsymbol{z}),\qquad[\rho(g)\cdot f](\boldsymbol{v},\boldsymbol{z})=f(\boldsymbol{v},g^{-1}\cdot\boldsymbol{z}),\tag{18}$$

它们定义了 $\mathcal{O}_d$ 在这些 L^2 空间上的标准左作用表示。

2.2 球谐函数与交织算子

球谐函数。次数为 ℓ 的球谐函数定义为限制在 $\mathbb{S}^{d-1}$ 上的次数为 ℓ 的齐次调和多项式，也就是说，满足 $P(t\boldsymbol{z})=t^\ell P(\boldsymbol{z})$ 且 $\Delta P=0$ 的多项式 P，其中 Δ 为拉普拉斯算子。我们用 $\mathscr{S}_{d,\ell}$ 表示这类函数构成的空间，并且 $N_{d,\ell}:=\dim(\mathscr{S}_{d,\ell})$，其中

$$N_{d,0}=1,\qquad N_{d,1}=d,\qquad N_{d,\ell}=\frac{d+2\ell-2}{\ell}\binom{d+\ell-3}{\ell-1}\quad\text{for }\ell\ge2.$$

特别 地，$|N_{d,\ell}/\binom{d}{\ell}-1|\le C_\ell/d$，其中 $C_\ell>0$ 是一个仅取决于 ℓ.

族 $\{\mathscr{S}_{d,\ell}\}_{\ell\ge0}$ 构成了 $L^2(\mathbb{S}^{d-1})$ 中关于内积 (16) 的一组相互正交子空间。这给出了正交分解

$$L^2(\mathbb{S}^{d-1})=\bigoplus_{\ell=0}^{\infty}\mathscr{S}_{d,\ell}.\tag{19}$$

每个子空间 $\mathscr{S}_{d,\ell}$ 在正交群的左作用（18）下都是不可约子空间$\mathcal{O}_d$。因此，（19）恰好是 L^2（$\mathbb{S}^{d-1}$）分解为不可约表示的半单（彼得-外尔）分解。

与无迹对称张量的同构。 存在一个 $\mathcal{O}_d$-等变等距映射，将 $\mathrm{TSym}_\ell(\mathbb{R}^d)$ 映射到 $\mathscr{S}_{d,\ell}$ 二者分别赋有各自的内积。对于任意 $\ell,\ d \in \mathbb{N}$，定义调和张量 $\mathcal{H}_{d,\ell} : \mathbb{S}^{d-1} \to \mathrm{TSym}_\ell(\mathbb{R}^d)$ 如下：[8]

$$\mathcal{H}_{d,\ell}(\boldsymbol{z}) = \kappa_{d,\ell}\sqrt{N_{d,\ell}}\mathsf{P}_{\mathsf{tf}}(\boldsymbol{z}^{\otimes \ell}), \qquad \kappa_{d,\ell} = \sqrt{2^\ell \frac{(d/2-1)_\ell}{(d-2)_\ell}} \tag{20}$$

其中，P_{tf} 是到无迹对称张量上的投影（见附录D.1中的显式表达式（86）），$(a)_\ell = a(a+1)\cdots(a+\ell-1)$ 是（上升）Pochhammer 符号。常数$\kappa_{d,\ell}$ 的选取使得 $\kappa_{d,\ell}\|\mathsf{P}_{\mathsf{tf}}(\boldsymbol{w}^{\otimes \ell})\|_{\mathsf{F}} = 1$ 对所有 $\|\boldsymbol{w}\|_2 = 1$ 都成立，并满足 $\kappa_{d,\ell} = 1 + \Theta_d(d^{-1/2})$。于是，

$$\Phi_{d,\ell} : \mathrm{TSym}_\ell(\mathbb{R}^d) \to \mathscr{S}_{d,\ell}, \qquad \Phi_{d,\ell}(\boldsymbol{A})(\boldsymbol{z}) = \langle \boldsymbol{A}, \mathcal{H}_{d,\ell}(\boldsymbol{z})\rangle_{\mathsf{F}} \tag{21}$$

是一个等距映射（见附录19中的引理D.3），也就是说，

$$\langle \Phi_{d,\ell}(\boldsymbol{A}), \Phi_{d,\ell}(\boldsymbol{B})\rangle_{L^2} = \langle \boldsymbol{A}, \boldsymbol{B}\rangle_{\mathsf{F}}, \qquad \text{for all } \boldsymbol{A}, \boldsymbol{B} \in \mathrm{TSym}_\ell(\mathbb{R}^d). \tag{22}$$

此外，尔，这一映射交织了 $\mathcal{O}_d$的张量表示与函数表示

$$\Phi_{d,\ell}(g \cdot \boldsymbol{A}) = \rho(g) \cdot \Phi_{d,\ell}(\boldsymbol{A}), \qquad \text{for all } \boldsymbol{A} \in \mathrm{TSym}_\ell(\mathbb{R}^d) \text{ and } g \in \mathcal{O}_d. \tag{23}$$

注意，$\mathcal{H}_{d,\ell}(\boldsymbol{z})$ 的所有元素都属于 $\mathscr{S}_{d,\ell}$，且（22）意味着

$$\mathbb{E}_{\tau_d}[\mathcal{H}_{d,\ell}(\boldsymbol{z}) \otimes \mathcal{H}_{d,\ell}(\boldsymbol{z})] = \mathrm{id}_{\mathrm{TSym}_\ell(\mathbb{R}^d)} \in \mathbb{L}_{d,\ell}.$$

调和分解。 结合（19）以及 $\Phi_{d,\ell}$ 是一个等距同构这一事实（见引理 19），任意 $f \in L^2(\mathbb{S}^{d-1})$ 都可以展开为调和张量之和（L^2）

$$f(\boldsymbol{z}) = \sum_{\ell=0}^{\infty} \langle \boldsymbol{A}_{d,\ell}, \mathcal{H}_{d,\ell}(\boldsymbol{z})\rangle_{\mathsf{F}}, \quad \text{where} \quad \boldsymbol{A}_{d,\ell} = \mathbb{E}_{\boldsymbol{z}\sim\tau_d}[f(\boldsymbol{z})\mathcal{H}_{d,\ell}(\boldsymbol{z})] \in \mathrm{TSym}_\ell(\mathbb{R}^d).$$

此外，若 f 对某个 $\boldsymbol{W} \in \mathrm{Stf}_{\mathsf{s}}(\mathbb{R}^d)$ 具有 $\mathcal{O}_d^{\boldsymbol{W}}$-不变性，即 f 仅依赖于 $\boldsymbol{W}^\mathsf{T}\boldsymbol{z}$，则可写为 $\boldsymbol{A}_{d,\ell} = \mathsf{P}_{\mathsf{tf}}(\boldsymbol{W}^{\otimes \ell}\boldsymbol{B}_{\mathsf{s},\ell})$，其中 $\boldsymbol{B}_{\mathsf{s},\ell} \in \mathrm{Sym}_\ell(\mathbb{R}^{\mathsf{s}})$（见附录 D.3 中的引理 20）。一个重要的例子对应于 $\mathsf{s} = 1$ 和纬向调和函数

$$\langle \boldsymbol{A}_{d,\ell}, \mathcal{H}_{d,\ell}(\boldsymbol{z})\rangle_{\mathsf{F}} = b_{1,\ell}\langle \mathsf{P}_{\mathsf{tf}}(\boldsymbol{w}^{\otimes \ell}), \mathcal{H}_{d,\ell}(\boldsymbol{z})\rangle_{\mathsf{F}} = b_{1,\ell}\kappa_{d,\ell}Q_\ell^{(d)}(\langle \boldsymbol{w}, \boldsymbol{z}\rangle),$$

其中 $Q_\ell^{(d)}:\ [-1,1] \to 1$ 是带有 $\mathbb{E}_{\tilde{\tau}_{d,\,1}}[Q_\ell^{(d)}(z_1)Q_\ell^{(d)}(z_1)] = \delta_{\ell\ell'}$ 的（归一化）Gegenbauer 多项式，其中 $z_1 \sim \tilde{\tau}_{d,\,1}$ 是 $\langle \boldsymbol{e}_1,\ \boldsymbol{z}\rangle$ 在 $\boldsymbol{z} \sim \tau_d$ 下的边缘分布。

在 $f \in L^2(\mathcal{V} \otimes \mathbb{S}^{d-1}) = \bigoplus_{\ell\geq 0} L^2(\mathcal{V}) \otimes \mathscr{S}_{d,\,\ell}$ 的情形下，我们将其分解为

$$f(\boldsymbol{v}, \boldsymbol{z}) = \sum_{\ell=0}^{\infty} \langle \boldsymbol{A}_{d,\ell}(\boldsymbol{v}), \mathcal{H}_{d,\ell}(\boldsymbol{z})\rangle_{\mathsf{F}}, \qquad \boldsymbol{A}_{d,\ell}(\boldsymbol{v}) = \mathbb{E}_{\boldsymbol{z}\sim\tau_d}[f(\boldsymbol{v}, \boldsymbol{z})\mathcal{H}_{d,\ell}(\boldsymbol{z})|\boldsymbol{v}] \in \mathrm{TSym}_\ell(\mathbb{R}^d), \tag{24}$$

其中 $\mathbb{E}_{\boldsymbol{v}\sim\nu}[\|\boldsymbol{A}_{d,\ell}(\boldsymbol{v})\|_{\mathsf{F}}^2] < \infty$。

[8] 请注意，这与 [JKMS25] 不同，但归一化方式更自然，其中 $\mathcal{H}_{d,\,\ell}(\boldsymbol{z}) = \kappa_{d,\,\ell}^2\sqrt{N_{d,\,\ell}}\mathsf{P}_{\mathsf{tf}}(\boldsymbol{z}^{\otimes \ell})$。

张量积表示。 我们还将考虑张量表示 $\mathrm{TSym}^p(\mathbb{R}^d)\otimes\mathrm{TSym}^q(\mathbb{R}^d)$，其作用为 $g\cdot(\boldsymbol{A}\otimes\boldsymbol{B}) = (g\cdot\boldsymbol{A})\otimes(g\cdot\boldsymbol{B})$，该表示允许半单分解

$$\mathrm{TSym}_p(\mathbb{R}^d)\otimes\mathrm{TSym}_q(\mathbb{R}^d)\cong\bigoplus_{j=0}^{\min(p,q)}\mathrm{TSym}_{p+q-2j}(\mathbb{R}^d). \tag{25}$$

特别地，我们有 $\mathscr{S}_{d,p}\otimes\mathscr{S}_{d,q}$ 的显式调和分解：对于任意 $\boldsymbol{C}\in\mathrm{TSym}^p(\mathbb{R}^d)\otimes\mathrm{TSym}^q(\mathbb{R}^d)$,

$$\langle\boldsymbol{C},\mathcal{H}_{d,p}(\boldsymbol{z})\otimes\mathcal{H}_{d,q}(\boldsymbol{z})\rangle_{\mathsf{F}} = \sum_{j=0}^{\min(p,q)} b^{(d)}_{p,q,j}\langle\mathsf{P}^{(p,q)}_{\mathsf{tf},j}(\boldsymbol{C}),\mathcal{H}_{d,p+q-2j}(\boldsymbol{z})\rangle_{\mathsf{F}},$$

其中 $b^{(d)}_{p,q,j} = \Theta_d(1)$ 与克莱布什-戈登系数有关，而 $\mathsf{P}^{(p,q)}_{\mathsf{tf},j}:\mathrm{TSym}^p(\mathbb{R}^d)\otimes\mathrm{TSym}^q(\mathbb{R}^d)\to\mathrm{TSym}^{p+q-2j}(\mathbb{R}^d)$ 是如下定义的线性算子：对所有 $\boldsymbol{A}\in\mathrm{TSym}^p(\mathbb{R}^d)$、$\boldsymbol{B}\in\mathrm{TSym}^q(\mathbb{R}^d)$,

$$\mathsf{P}^{(p,q)}_{\mathsf{tf},j}(\boldsymbol{A}\otimes\boldsymbol{B}) = \boldsymbol{A}\diamond_j\boldsymbol{B} := \mathsf{P}_{\mathsf{tf}}(\mathsf{P}_{\mathsf{sym}}(\boldsymbol{A}\otimes_j\boldsymbol{B})) \tag{26}$$

当 $0\le j\le\min(p,q)$ 时，而在其他情况下 $\mathsf{P}^{(p,q)}_{\mathsf{tf},j}(\boldsymbol{A}\otimes\boldsymbol{B})$。

舒尔正交关系与超收缩性。 每个 $\mathscr{S}_{d,\ell}$（等价地，$\mathrm{TSym}^\ell(\mathbb{R}^d)$）都构成 $\mathcal{O}_d$ 的一个不可约酉表示。函数

$$g\mapsto\langle\rho(g)\cdot f,h\rangle_{L^2(\tau_d)},\qquad f,h\in\mathscr{S}_{d,\ell}, \tag{27}$$

是该表示的矩阵系数，并满足舒尔正交关系：对于任意 $f,h\in\mathscr{S}_{d,\ell}$ 和 $f',h'\in\mathscr{S}_{d,k}$,

$$\int_{\mathcal{O}_d}\langle\rho(g)\cdot f,h\rangle_{L^2}\,\langle\rho(g)\cdot f',h'\rangle_{L^2}\,\pi_d(\mathrm{d}g) = \frac{\delta_{\ell k}}{N_{d,\ell}}\,\langle f,f'\rangle_{L^2}\,\langle h,h'\rangle_{L^2}. \tag{28}$$

根据酉等价，当 f,h 替换为 $\mathrm{TSym}^\ell(\mathbb{R}^d)$ 中的无迹对称张量时，同一关系仍成立（我们有等式 $\langle\rho(g)\cdot\Phi_{d,\ell}(\boldsymbol{A})\Phi_{d,\ell}(\boldsymbol{B})\rangle_{L^2} = \langle g\cdot\boldsymbol{A},\boldsymbol{B}\rangle_{\mathsf{F}}$）。

记 $\mathcal{M}_{d,\ell}\subset L^2(\mathcal{O}_d)$ 为所有矩阵系数（27）的张成空间，其中 $f,h\in\mathscr{S}_{d,\ell}$。利用 Gross 定理、连通分支 $\mathcal{SO}_d$上的热半群，以及与 $\mathscr{S}_{d,\ell}$相关联的卡西米尔特征值，子空间 $\mathcal{M}_{d,\ell}$ 满足超收缩性性质（见附录D.5）：对于任意 $d\ge3$、任意 $F\in\mathcal{M}_{d,\ell}$以及任意 $1<p\le q<\infty$,

$$\|F\|_{L^q(\mathcal{O}_d)}\le 2^{\frac{1}{p}-\frac{1}{q}}\left(\frac{q-1}{p-1}\right)^{\gamma_d(\ell)}\|F\|_{L^p(\mathcal{O}_d)},\qquad\gamma_d(\ell)=\frac{2\ell(\ell+d-2)}{d-2}. \tag{29}$$

我们还要回顾关于球面 [Bec92] 上多项式超收缩性的以下著名结果：

$$\|f\|_{L^q(\tau_d)}\le\left(\frac{q-1}{p-1}\right)^{\ell/2}\|f\|_{L^p(\tau_d)}, \tag{30}$$

对于所有次数至多为 ℓ的多项式 f。

3 下界与复杂度飞跃

我们首先为球面多指标模型的学习建立计算下界。正如引言中所述，在温和的假设下，这些模型在信息论意义上可学习，所需样本数为 $\Theta_d(d)$；例如参见 [JKMS25，附录 H]。然而，有人猜想，对于一般的链接函数，任何多项式时间算法都无法利用这么多样本恢复潜在子空间 ν_d [BKM+19、DPVLB24、DLB25、DIKR25、JKMS25]。也就是说，这些模型呈现出所谓的统计-计算差距。要排除所有多项式时间算法，就需要解决NP $\neq$P问题，因此标准做法是在受限但强大的计算模型中证明其困难性。本文采用流行的统计查询 (SQ) 和低阶多项式 (LDP) 框架，为球面 MIM 推导此类计算下界。

3.1 统计查询与低次多项式框架

我们简要回顾统计查询和局部差分隐私框架，并建议参阅 [Fel17、Rey20、Wei25、Mis25] 以获取更全面的介绍。这两个框架都将估计问题转化为一个检测问题。

假设我们观察到的数据是从一族分布 $\{\mathbb{P}_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ 中生成的。我们通过更简单的任务——区分——的复杂度，来给估计 $\boldsymbol{\theta}$ （使用 n 个样本）的复杂度给出下界

$$H_0 : (\boldsymbol{y}_i, \boldsymbol{z}_i)_{i\in[n]} \overset{\text{i.i.d.}}{\sim} \mathbb{P}_0 \qquad \text{versus} \qquad H_1 : (\boldsymbol{y}_i, \boldsymbol{z}_i)_{i\in[n]} \overset{\text{i.i.d.}}{\sim} \mathbb{P}_{\boldsymbol{\theta}} \text{ for some } \boldsymbol{\theta} \in \boldsymbol{\Theta}. \tag{31}$$

对于某个零假设 $\mathbb{P}_0$。尽管估计可能比检测严格更难，检测下界仍提供了稳健的基准。

统计查询算法。 统计查询框架由 Kearns 引入 [Kea98]，该框架对算法进行建模：算法仅通过对查询函数期望值的带噪估计与数据交互，而不直接访问样本。广泛的一类算法——包括基于梯度的方法、支持向量机和马尔可夫链蒙特卡罗——都可以在此模型中实现。统计查询框架通常被认为能够刻画抗噪算法的能力。

对于查询次数 q 和容差 $\tau_{\mathrm{SQ}} > 0$，一个统计查询算法 $\mathcal{A} \in$ 统计查询(q, τ_{SQ}) 针对检测任务（31）接收一个输入分布 $\mathbb{P} \in \{\mathbb{P}_0\} \cup \{\mathbb{P}_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ 并分 q 轮进行。在每一轮 $t \in \{1, \dots, q\}$，它发出一个查询 $\phi_t : \mathcal{Y} \times \mathcal{Z} \to [-1, 1]$ 并接收一个响应 v_t满足

$$|v_t - \mathbb{E}_{(\boldsymbol{y},\boldsymbol{z})\sim\mathbb{P}}[\phi(\boldsymbol{y}, \boldsymbol{z})]| \le \tau_{\mathrm{SQ}}. \tag{32}$$

ϕ_t的选择可以依赖于先前的响应 v_1，……，v_{t-1}。经过 q轮后，该算法输出0或1。我们称 $\mathcal{A}$ 成功区分 $\mathbb{P}_0$ 与 $\{\mathbb{P}_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$，如果对于任意有效响应 v_t，$\mathcal{A}$ 在 $\mathbb{P} = \mathbb{P}_0$时输出0，并在 $\mathbb{P} = \mathbb{P}_{\boldsymbol{\theta}}$ 时输出1，其中针对某个 $\boldsymbol{\theta} \in \Omega$。

预言机响应（32）可以使用 $n \asymp 1/\sqrt{\tau_{\mathrm{SQ}}}$ 个样本，通过经验平均来实现。因此，查询复杂度 q/τ_{SQ}^2 常被用作该算法运行时间的代理下界，其启发式假设是模拟每个查询需要$\Omega(1/\tau_{\mathrm{SQ}}^2)$的计算量。

低次多项式算法。 低次多项式框架是研究高维推断中统计—计算权衡的一种互补方法。从概念上讲，它刻画了这样一类算法：其检验统计量可以表示为观测数据中次数有界的多元多项式。

设 $\boldsymbol{\theta} \sim \pi$ 为参数空间上的一个先验，并考虑似然比

$$\mathcal{R}((\boldsymbol{y}_i, \boldsymbol{z}_i)_{i\in[n]}) = \mathbb{E}_{\boldsymbol{\theta}\sim\pi}\left[\prod_{i\in[n]} \frac{\mathrm{d}\mathbb{P}_{\boldsymbol{\theta}}}{\mathrm{d}\mathbb{P}_0}(\boldsymbol{y}_i, \boldsymbol{z}_i)\right].$$

令 $\mathsf{P}_{\le D}$ 表示 $L^2(\mathbb{P}_0^{\otimes n})$ 中投影到协变量 $\{\boldsymbol{z}_i\}_{i\in[n]}$ 上次数至多为 D 的多项式子空间的正交投影。低次似然比定义为

$$\mathcal{R}_{\le D}((\boldsymbol{y}_i, \boldsymbol{z}_i)_{i\in[n]}) = \mathsf{P}_{\le D}\mathcal{R}((\boldsymbol{y}_i, \boldsymbol{z}_i)_{i\in[n]}).$$

次数 D 被视为计算复杂度的一个替代指标：在 $O(n)$ 个变量中计算次数为 D 的多项式，所需时间至多为 $n^{O(D)}$。

低次猜想 [Hop18] 断言，若 $D = \omega_n(\log n)$，$\|\mathcal{R}_{\le D}\|_{L^2}^2 = 1 + o_n(1)$，则不存在能够实现弱检测的多项式时间算法（即相较于随机猜测具有非消失优势）。换言之，如果所有次数为$\omega_n(\log n)$ 的多项式检验均失败，这就被视为强有力的证据，表明针对问题 [Wei25]，所有多项式时间算法都会失败。

3.2 弱恢复下界：对齐复杂度

我们首先为弱恢复任务建立计算下界，即以优于随机猜测的精度估计 $\boldsymbol{W}_*$。为刻画这一概念，我们考虑检测问题

$$H_0 : (\boldsymbol{y}_i, \boldsymbol{z}_i)_{i\in[n]} \overset{\text{i.i.d.}}{\sim} \mathbb{P}_{\nu_d,\emptyset} \quad \text{v.s.} \quad H_1 : (\boldsymbol{y}_i, \boldsymbol{z}_i)_{i\in[n]} \overset{\text{i.i.d.}}{\sim} \mathbb{P}_{\nu_d}^{\boldsymbol{W}} \text{ for some } \boldsymbol{W} \in \mathrm{Stf}_{\mathsf{s}}(\mathbb{R}^d), \tag{33}$$

其中零分布$\mathbb{P}_{\nu_d,\emptyset} = \nu_d^Y \otimes \tau_d$ 对应于响应 $\boldsymbol{y}$ 与输入 $\boldsymbol{z}$相互独立的模型。我们还记$\nu_{d,\emptyset} := \nu_d^Y \otimes \tilde{\tau}_{d,\mathsf{s}}$。

对$\boldsymbol{W} \sim \pi_{d,\mathsf{s}}$ 取平均，并利用 τ_d的旋转不变性，我们得到

$$\mathbb{P}_{\nu_d,\emptyset} = \mathbb{E}_{\boldsymbol{W}\sim\pi_{d,\mathsf{s}}}\left[\mathbb{P}_{\nu_d}^{\boldsymbol{W}}\right] = \mathbb{E}_{g\sim\pi_d}\left[\mathbb{P}_{\nu_d}^{g\cdot\boldsymbol{W}_*}\right] = \mathbb{E}_{g\sim\pi_d}\left[\rho(g)\cdot\mathbb{P}_{\nu_d}^{\boldsymbol{W}_*}\right],$$

并且$\mathbb{P}_{\nu_d,\emptyset}$ 恰好是在 $\boldsymbol{W}$上的均匀先验下植入模型的混合。

为应用局部差分隐私框架，我们假设该模型相对于零分布具有有限的卡方散度（见注3.2）：

假设 1. 我们有$\nu_d \ll \nu_{d,\emptyset}$，并且拉东-尼科迪姆导数满足$\frac{\mathrm{d}\nu_d}{\mathrm{d}\nu_{d,\emptyset}} \in L^2(\nu_{d,\emptyset})$。

等价地，似然比

$$\frac{\mathrm{d}\mathbb{P}_{\nu_d}^{\boldsymbol{W}}}{\mathrm{d}\mathbb{P}_{\nu_d,\emptyset}} \in L^2(\mathbb{P}_{\nu_d,\emptyset})$$

具有有限的 L^2-范数。此范数与 $\boldsymbol{W} \in \mathrm{Stf}_{\mathsf{s}}(\mathbb{R}^d)$ 的具体选取无关。利用分解（24），我们将似然比展开为

$$\frac{\mathrm{d}\mathbb{P}_{\nu_d}^{\boldsymbol{W}}}{\mathrm{d}\mathbb{P}_{\nu_d,\emptyset}}(\boldsymbol{y}, \boldsymbol{z}) = 1 + \sum_{\ell=1}^{\infty} \langle \boldsymbol{\xi}_{\emptyset,\ell}^{\boldsymbol{W}}(\boldsymbol{y}), \mathcal{H}_{d,\ell}(\boldsymbol{z})\rangle_{\mathsf{F}}, \qquad \boldsymbol{\xi}_{\emptyset,\ell}^{\boldsymbol{W}}(\boldsymbol{y}) := \mathbb{E}_{\mathbb{P}_{\nu_d}^{\boldsymbol{W}}}\left[\mathcal{H}_{d,\ell}(\boldsymbol{z}) \mid \boldsymbol{y}\right].$$

对于每个次数 $\ell \ge 1$，定义二阶矩算子

$$\boldsymbol{\Gamma}_{\emptyset,\ell}^{\boldsymbol{W}} := \mathbb{E}_{\nu_{d,\emptyset}}[\boldsymbol{\xi}_{\emptyset,\ell}^{\boldsymbol{W}}(\boldsymbol{y}) \otimes \boldsymbol{\xi}_{\emptyset,\ell}^{\boldsymbol{W}}(\boldsymbol{y})] \in \mathbb{L}_{d,\ell}.$$

我们通常省略上标 $\boldsymbol{W}$，直接写作 $\boldsymbol{\xi}_{\emptyset,\ell}$ 和 $\boldsymbol{\Gamma}_{\emptyset,\ell}$。特别地，$\mathrm{Tr}(\boldsymbol{\Gamma}_{\emptyset,\ell})$、$\|\boldsymbol{\Gamma}_{\emptyset,\ell}\|_{\mathrm{op}}$和$\|\boldsymbol{\Gamma}_{\emptyset,\ell}\|_{\mathsf{F}}$ 均不依赖于 $\boldsymbol{W}$。

对于弱恢复，我们建立了两个互补的下界：(i) 统计查询框架中的查询复杂度下界（可作为算法运行时间的替代指标），以及 (ii) 低次多项式框架中的样本复杂度下界（对所有多项式时间算法构成障碍）。它们分别由查询对齐 复杂度和样本对齐 复杂度刻画：

$$\mathsf{qAlign}(\nu_d\|\overline{\nu}_{d,\emptyset}) = \inf_{\ell\geq 1} \frac{N_{d,\ell}}{\|\boldsymbol{\Gamma}_{\emptyset,\ell}\|_{\mathrm{op}}}, \qquad \mathsf{sAlign}(\nu_d\|\overline{\nu}_{d,\emptyset}) = \inf_{\ell\geq 1} \frac{\sqrt{N_{d,\ell}}}{\|\boldsymbol{\Gamma}_{\emptyset,\ell}\|_{\mathsf{F}}}. \tag{34}$$

另请注意，$\mathrm{d}\mathbb{P}^{\boldsymbol{W}}_{\nu_d}/\mathrm{d}\mathbb{P}_{\nu_d,\emptyset}$ 是 $\mathcal{O}_d^{\boldsymbol{W}}$不变的。因此，根据引理 20 （附录D），我们可以将系数写为 $\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}) = \mathsf{P}_{\mathsf{tf}}(\boldsymbol{W}^{\otimes\ell}\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y}))$. 特别地，$\boldsymbol{\xi}_{\emptyset,\ell}$的秩至多为 s^ℓ，并且

$$\|\boldsymbol{\Gamma}_{\emptyset,\ell}\|_{\mathrm{op}} = \Theta_d(\|\boldsymbol{\xi}_{\emptyset,\ell}\|^2_{L^2}), \quad \|\boldsymbol{\Gamma}_{\emptyset,\ell}\|_{\mathsf{F}} = \Theta_d(\|\boldsymbol{\xi}_{\emptyset,\ell}\|^2_{L^2}), \quad \text{where } \|\boldsymbol{\xi}_{\emptyset,\ell}\|^2_{L^2} := \mathrm{Tr}(\boldsymbol{\Gamma}_{\emptyset,\ell}) = \mathbb{E}[\|\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y})\|^2_{\mathsf{F}}].$$

我们总有 $\|\boldsymbol{\Gamma}_{\emptyset,\ell}\|_{\mathrm{op}} \leq 1$ 且 $\|\boldsymbol{\Gamma}_{\emptyset,\ell}\|_{\mathsf{F}} \leq \mathsf{s}^{\ell/2}$。

定理1 （球面 MIM 的弱恢复下界）。

(a) （查询下界。）设 ν_d 是一个满足假设*1*的球面多指标模型。如果算法$\mathcal{A} \in$ 统计查询(q,τ_{SQ}) 成功地区分 $\{\mathbb{P}^{\boldsymbol{W}}_{\nu_d}: \ \boldsymbol{W} \in \mathrm{Stfs}(\mathbb{R}^d)\}$ 与 $\mathbb{P}_{\nu_d,\emptyset}$，那么

$$q/\tau^2_{\mathrm{SQ}} \geq \mathsf{qAlign}(\nu_d\|\overline{\nu}_{d,\emptyset}).$$

(b) （样本下界。）设 $\{\nu_d\}_{d\geq 1}$ 是满足假设*1*的球面 *MIM* 序列。假设存在 $p \in \mathbb{N}$，使得 $\mathsf{sAlign}(\nu_d\|\nu_{d,\emptyset}) = O_d(d^{p/2})$。若对于 $D = o_d(d^{17(p+2)})$，

$$n = o_d\left(\frac{\mathsf{sAlign}(\nu_d\|\overline{\nu}_{d,\emptyset})}{D^{9p/2-1}}\right), \tag{35}$$

则 $\|\mathcal{R}_{\leq D}\|_{L^2} = 1 + o_d(1)$，并且在低次猜想成立的条件下，没有多项式时间算法能够在 $\{\mathbb{P}^{\boldsymbol{W}}_{\nu_d}: \ \boldsymbol{W} \in \mathrm{Stfs}(\mathbb{R}^d)\}$ 与 $\mathbb{P}_{\nu_d,\emptyset}$ 之间实现弱检测。

这些下界是通过应用舒尔正交关系（28）、标准的二阶矩论证（针对统计查询）以及矩阵系数的超收缩性（29）（针对局部差分隐私）得到的。定理1 的详细证明见附录A。

定理1中的下界在不可约子空间$\mathscr{S}_{d,\ell}$ 上彼此解耦，并且有一种简单的启发式解释。固定 $\ell \geq 1$，并考虑其统计量仅依赖于 z的次数为ℓ 的球谐函数的估计量。对于此类估计量，下界可化为$q/\tau^2_{\mathrm{SQ}} \gtrsim N_{d,\ell}/\|\boldsymbol{\Gamma}_{\emptyset,\ell}\|_{\mathrm{op}}$ 以及$n \gtrsim \sqrt{N_{d,\ell}}/\|\boldsymbol{\Gamma}_{\emptyset,\ell}\|_{\mathsf{F}}$ （详见 [JKMS25，附录 C]）。随后，对 $\ell \geq 1$取下确界（即最佳复杂度），即可得到全局界（34）。因此，这些下界将问题分解为彼此独立的检测子问题，每个问题对应一个调和子空间 $\mathscr{S}_{d,\ell}$；式（34）中的每一项都对应于受限于次数为ℓ的球谐函数的算法的复杂度。

在第4节中，我们将针对每个子空间$\mathscr{S}_{d,\ell}$构造（几乎）达到这些下界的算法。这表明，应选择在 ℓ中实现最小值的次数qAlign （对于运行时间最优程序），或在sAlign中实现最小值的次数（对于样本最优程序）；参见公式（5）。

注3.1 （检测-恢复差距）。 定理1 中的界针对的是检测任务（33），而非直接针对估计。在（34）中的下确界由 $\ell = 1$ 和 $\sqrt{d}/\|\boldsymbol{\Gamma}_{\emptyset,1}\|_{\mathsf{F}} \ll d$ 处取得时，会出现检测-恢复差距。在这种情况下，这些界对检测是紧的，但从信息论意义上讲，估计仍需要 $n = \Omega_d(d)$ 个样本。此外，任何算法仅为处理这些 n样本就至少需要 $\Omega(dn)$ 的运行时间。

注3.2 （卡方散度有限）。 假设 1 要求似然比平方可积。这是应用低次多项式框架的必要条件，也可理解为假设标签中存在足够的噪声。在实践中，可以通过向响应添加少量高斯噪声来满足这一条件。对于统计查询下界，可以通过类似 [DIKR25] 中的论证放宽这一假设。更广泛地说，假设1排除了在无噪声设置下表现优于这些界（34）的非稳健算法（例如，参见 [SZB21]）。还需注意，$\mathbf{\Gamma}_{\emptyset,\ell}$ 总是有定义的（其中 $\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}) := \mathbb{E}_{\mathbb{P}_{\nu_d}}[\mathcal{H}_{d,\ell}(\boldsymbol{z})|\boldsymbol{y}]$），即使卡方散度为无穷大时也是如此；对齐复杂度（以及跃迁复杂度）也总是有定义的。

注 3.3 （额外的权衡）。 局部差分隐私下界（35）为样本复杂度中$\tilde{}$量级的潜在权衡 $D^{-\Theta(1)}$ 留下了空间，这可以通过考虑次数为D 的算法（其运行时间为 $d^{\Theta(D)}$）来实现。在张量主成分分析中，已知这类权衡是紧的（在低次猜想成立的前提下） [WEAM19、KX26、Li25]。探索当前设定下类似的权衡是一个有趣的方向，我们将其留待未来研究。

3.3 强恢复的下界：跃迁复杂度

现在我们转向强恢复这一任务，即恢复完整子空间 $\boldsymbol{W}_*$。正如引言中所讨论的，最优方法将分多个步骤进行，逐步恢复 $\boldsymbol{W}_*$。我们通过考虑一族部分检测问题来刻画这一过程的复杂度，这些问题对应于一系列中间阶段，在这些阶段中， $\boldsymbol{W}_*$ 的一个真子空间仅有一部分已被恢复。

为简化起见，且不失一般性，我们固定一个参考标架 $\boldsymbol{W}_* \in \mathsf{Stf}_{\mathsf{s}}(\mathbb{R}^d)$。定义 $\mathsf{Sub}(\boldsymbol{W}_*)$ 为 $\boldsymbol{W}_*$ 的所有严格子框架的集合，即所有满足 $0 \le r < \mathsf{s}$ 且 $\boldsymbol{U} \subset \boldsymbol{W}_*$（也就是说，$\mathsf{span}(\boldsymbol{U}) \subset \mathsf{span}(\boldsymbol{W}_*)$）的 $\boldsymbol{U} \in \mathsf{Stf}_r(\mathbb{R}^d)$。记 $s_{\boldsymbol{U}} := \mathsf{rank}(\boldsymbol{U})$ 和 $d_{\boldsymbol{U}} := d - s_{\boldsymbol{U}}$，并令 $\boldsymbol{U}_\perp \in \mathsf{Stf}_{d_{\boldsymbol{U}}}(\mathbb{R}^d)$ 为 $\boldsymbol{U}$ 在 $\mathbb{R}^d$中的一个正交补。

考虑数据 $(\boldsymbol{y}, \boldsymbol{z}) \sim \mathbb{P}_{\nu_d}^{\boldsymbol{W}_*}$，并假设我们已经恢复了一个子空间 $\boldsymbol{U} \in \mathsf{Sub}(\boldsymbol{W}_*)$。我们将输入分解为

$$\boldsymbol{z} = \boldsymbol{U}\boldsymbol{r}_{\boldsymbol{U}} + \sqrt{1 - \|\boldsymbol{r}_{\boldsymbol{U}}\|_2^2}\, \boldsymbol{U}_\perp \boldsymbol{z}_{\boldsymbol{U}}, \qquad \boldsymbol{r}_{\boldsymbol{U}} := \boldsymbol{U}^\mathsf{T}\boldsymbol{z} \in \mathbb{R}^{s_{\boldsymbol{U}}}, \qquad \boldsymbol{z}_{\boldsymbol{U}} = \frac{\boldsymbol{U}_\perp^\mathsf{T}\boldsymbol{z}}{\sqrt{1 - \|\boldsymbol{r}_{\boldsymbol{U}}\|_2^2}} \in \mathbb{S}^{d_{\boldsymbol{U}}-1}, \quad (36)$$

使得 $(\boldsymbol{r}_{\boldsymbol{U}}, \boldsymbol{z}_{\boldsymbol{U}}) \sim \tilde{\tau}_{d,s_{\boldsymbol{U}}} \otimes \tau_{d_{\boldsymbol{U}}}$。接着，我们通过将新的响应定义为 $\boldsymbol{y}_{\boldsymbol{U}} := (\boldsymbol{y}, \boldsymbol{r}_{\boldsymbol{U}})$，其条件分布为，来定义一个新的球面多指标模型 $(\boldsymbol{y}_{\boldsymbol{U}}, \boldsymbol{z}_{\boldsymbol{U}}) \sim \mathbb{P}_{\nu_{d,\boldsymbol{U}}}$

$$\boldsymbol{y}_{\boldsymbol{U}}|\boldsymbol{z}_{\boldsymbol{U}} \sim \nu_{d,\boldsymbol{U}}(\mathrm{d}\boldsymbol{y}_{\boldsymbol{U}}|\boldsymbol{W}_*^\mathsf{T}\boldsymbol{U}_\perp\boldsymbol{z}_{\boldsymbol{U}}) = \nu_d\big(\mathrm{d}\boldsymbol{y}\big|\boldsymbol{W}_*^\mathsf{T}\boldsymbol{z}\big)\tilde{\tau}_{d,s_{\boldsymbol{U}}}(\mathrm{d}\boldsymbol{r}_{\boldsymbol{U}}). \quad (37)$$

我们将对应的部分零假设定义为

$$\mathbb{P}_{\nu_d,\boldsymbol{U}} = \nu_{d,\boldsymbol{U}}^Y \otimes \tau_{d_{\boldsymbol{U}}} = \mathbb{E}_{g\sim\mathrm{Unif}(\mathcal{O}_d^{\boldsymbol{U}})}\left[\mathbb{P}_{\nu_d}^{g\cdot\boldsymbol{W}_*}\right],$$

这里回顾一下， $\mathcal{O}_d^{\boldsymbol{U}}$对应于 $\mathcal{O}_d$ 中使 $\boldsymbol{U}$ 保持不变的子群。我们还记作$\nu_{d,\boldsymbol{U}} := \nu_{d\,,\boldsymbol{U}}^Y \otimes \tilde{\tau}_{d,\mathsf{s}-s_{\boldsymbol{U}}}$，其中$\nu_{d\,,\boldsymbol{U}}^Y$是 $(\boldsymbol{y}, \boldsymbol{U}^\mathsf{T}\boldsymbol{z})$ 的边缘分布。

接着，我们考虑部分检测问题

$$H_0 : (\boldsymbol{y}_{\boldsymbol{U},i}, \boldsymbol{z}_{\boldsymbol{U},i})_{i\in[n]} \overset{\text{i.i.d.}}{\sim} \mathbb{P}_{\nu_d,\boldsymbol{U}} \quad \text{v.s.} \quad H_1 : (\boldsymbol{y}_{\boldsymbol{U},i}, \boldsymbol{z}_{\boldsymbol{U},i})_{i\in[n]} \overset{\text{i.i.d.}}{\sim} \mathbb{P}_{\nu_d}^{g\cdot\boldsymbol{W}} \text{ for some } g \in \mathcal{O}_d^{\boldsymbol{U}}. \quad (38)$$

在我们的下界证明中，要求算法对由$\boldsymbol{U} \in \mathsf{Sub}(\boldsymbol{W}_*)$所指标化的所有此类部分问题都能成功。等价地，我们考虑最坏情形下的 $\boldsymbol{U}$ ，在这种情况下仍必须判断$\boldsymbol{y}_{\boldsymbol{U}}$ 和 $\boldsymbol{z}_{\boldsymbol{U}}$ 是否独立。

和之前一样, 我们现在要求相对于所有部分零假设都满足有限卡方条件 :

假设 2。 对于每个 $\boldsymbol{U} \in \mathrm{Sub}(\boldsymbol{W}_*)$，拉东-尼科迪姆导数满足 $\mathrm{d}\nu_{d,\boldsymbol{U}} / \mathrm{d}\nu_d \in L^2(\nu_d, \boldsymbol{U})$。

对于 $\boldsymbol{U} \subset \boldsymbol{W}$，我们可以像之前一样展开似然比

$$\frac{\mathrm{d}\mathbb{P}^{\boldsymbol{W}}_{\nu_d}}{\mathrm{d}\mathbb{P}_{\nu_d,\boldsymbol{U}}}(\boldsymbol{y}_{\boldsymbol{U}}, \boldsymbol{z}_{\boldsymbol{U}}) = 1 + \sum_{\ell=1}^{\infty} \langle \boldsymbol{\xi}^{\boldsymbol{W}}_{\boldsymbol{U},\ell}(\boldsymbol{y}), \mathcal{H}_{d_{\boldsymbol{U}},\ell}(\boldsymbol{z}_{\boldsymbol{U}}) \rangle_{\mathsf{F}}, \qquad \boldsymbol{\xi}^{\boldsymbol{W}}_{\boldsymbol{U},\ell}(\boldsymbol{y}) := \mathbb{E}_{\mathbb{P}^{\boldsymbol{W}}_{\nu_d}} \left[\mathcal{H}_{d_{\boldsymbol{U}},\ell}(\boldsymbol{z}_{\boldsymbol{U}}) \mid \boldsymbol{y}_{\boldsymbol{U}} \right],$$

并定义二阶矩算子

$$\boldsymbol{\Gamma}^{\boldsymbol{W}}_{\boldsymbol{U},\ell} := \mathbb{E}_{\nu_{d,\boldsymbol{U}}}[\boldsymbol{\xi}^{\boldsymbol{W}}_{\boldsymbol{U},\ell}(\boldsymbol{y}_{\boldsymbol{U}}) \otimes \boldsymbol{\xi}^{\boldsymbol{W}}_{\boldsymbol{U},\ell}(\boldsymbol{y}_{\boldsymbol{U}})] \in \mathbb{L}_{d_{\boldsymbol{U}},\ell}.$$

我们再次省略上标 $\boldsymbol{W}$，并记为 $\boldsymbol{\xi}_{\boldsymbol{U},\ell}$ 和 $\boldsymbol{\Gamma}_{\boldsymbol{U},\ell}$。特别地，$\mathrm{Tr}(\boldsymbol{\Gamma}_{\boldsymbol{U},\ell})$、$\|\boldsymbol{\Gamma}_{\boldsymbol{U},\ell}\|_{\mathrm{op}}$和$\|\boldsymbol{\Gamma}_{\boldsymbol{U},\ell}\|_{\mathsf{F}}$ 均不依赖于补空间 $\boldsymbol{W}\boldsymbol{U}_\perp$ 的选取。

对于强恢复，我们在所有由 38 所索引的部分问题（$\boldsymbol{U} \in \mathrm{Sub}(\boldsymbol{W}_*)$）中，选取最坏（即最困难）的弱恢复实例。由于每个这样的实例本身都是约化球面 MIM（37）的弱恢复问题，我们可以对每个 $\boldsymbol{U}$ 应用定理 查询跳跃和样本跳跃复杂度的定义如下：

$$\mathsf{qLeap}(\nu_d) = \sup_{\boldsymbol{U} \in \mathrm{Sub}(\boldsymbol{W}_*)} \mathsf{qAlign}(\nu_d \| \overline{\nu}_{d,\boldsymbol{U}}) = \sup_{\boldsymbol{U} \in \mathrm{Sub}(\boldsymbol{W}_*)} \inf_{\ell \geq 1} \frac{N_{d_{\boldsymbol{U}},\ell}}{\|\boldsymbol{\Gamma}_{\boldsymbol{U},\ell}\|_{\mathrm{op}}},$$

$$\mathsf{sLeap}(\nu_d) = \sup_{\boldsymbol{U} \in \mathrm{Sub}(\boldsymbol{W}_*)} \mathsf{sAlign}(\nu_d \| \overline{\nu}_{d,\boldsymbol{U}}) = \sup_{\boldsymbol{U} \in \mathrm{Sub}(\boldsymbol{W}_*)} \inf_{\ell \geq 1} \frac{\sqrt{N_{d_{\boldsymbol{U}},\ell}}}{\|\boldsymbol{\Gamma}_{\boldsymbol{U},\ell}\|_{\mathsf{F}}}.$$

下面的定理直接由定理 1 得出。

定理2（球面多指标模型强恢复的下界）。

(a)（查询下界。）设 ν_d 是满足假设2的球面多指标模型。若算法$\mathcal{A} \in$ 统计查询(q, τ_{SQ}) 成功完成强恢复任务(38)，则

$$q/\tau^2_{\mathrm{SQ}} \geq \mathsf{qLeap}(\nu_d).$$

(b)（样本下界。）设 $\{\nu_d\}_{d \geq 1}$ 是满足假设2的一列球面多指标模型。假设存在 $p \in \mathbb{N}$ 使得$\mathsf{sLeap}(\nu_d) = O_d(d^{p/2})$。若对于 $D = o_d(d^{1/(p+2)})$，

$$n = o_d\left(\frac{\mathsf{sLeap}(\nu_d)}{D^{9p/2-1}} \right),$$

则 $\|\mathcal{R}_{\leq D}\|_{L^2} = 1 + o_d(1)$ 并且在低次猜想成立的条件下，不存在多项式时间算法能够成功完成强恢复任务(38)。

上述跃迁复杂度促使我们采用一种自然的逐步恢复过程。固定一列调和次数 $\ell_1, \ell_2, \ldots \in \mathbb{N}$。第一步，我们应用一个基于次数-$\ell_1$ 球谐函数的估计量。假设该估计量恢复了一个 s_1 维子空间 $\boldsymbol{U}_1 \in \mathrm{Stf}_{s_1}(\mathbb{R}^d)$，满足 $\boldsymbol{U}_1 \subset \boldsymbol{W}_*$。第二步，我们改用次数-$\ell_2$ 调和估计量，此时在正交补 $\boldsymbol{U}_{1,\perp}$中的输入 $\boldsymbol{z}_{\boldsymbol{U}_1}$ 上进行。这会得到一组额外的新方向$\boldsymbol{U}_2 \subset \boldsymbol{U}_{1,\perp}$（满足$\boldsymbol{U}_2 \in \mathrm{Stf}_{s_2}(\mathbb{R}^{d-s_1})$），对应于 $\boldsymbol{W}_*$ 中与 $\boldsymbol{U}_1$ 正交的新分量。重复这一过程，我们得到一列嵌套子空间

$$\boldsymbol{U}_{\leq t} := \boldsymbol{U}_1 \oplus \boldsymbol{U}_2 \oplus \cdots \oplus \boldsymbol{U}_t \in \mathrm{Stf}_{s_{\leq t}}(\mathbb{R}^d) \subseteq \boldsymbol{W}_*, \qquad s_{\leq t} := s_1 + \cdots + s_t,$$

按照约定 $s_0 := 0$ 且 $\boldsymbol{U}_0 = \emptyset$。由于每个阶段至少必须恢复一个新方向，因此该过程会在若干步 $T \leq$ s 后终止，此时 $\boldsymbol{U}_{\leq T} = \boldsymbol{W}_*$。

该顺序过程的总成本由最难的 中间步骤决定。在步骤 t，相关的统计查询和局部差分隐私复杂度分别为$d^{\ell_t}/\|\boldsymbol{\Gamma}_{\boldsymbol{U}_{\leq t},\ \ell_t}\|_{\mathrm{op}}$ 和 $d^{\ell_t/2}/\|\boldsymbol{\Gamma}_{\boldsymbol{U}_{\leq t},\ \ell_t}\|_{\mathsf{F}}$。因此，总顺序复杂度为

$$\max_{t\in[T-1]} \frac{d^{\ell_t}}{\|\boldsymbol{\Gamma}_{\boldsymbol{U}_{\leq t},\ell_t}\|_{\mathrm{op}}}, \qquad \max_{t\in[T-1]} \frac{d^{\ell_t/2}}{\|\boldsymbol{\Gamma}_{\boldsymbol{U}_{\leq t},\ell_t}\|_{\mathsf{F}}}$$

分别对应统计查询和局部差分隐私设置。

运行时间最优 和样本最优 策略通过最小化相应的中间成本，归纳地选择调和次数：

$$\ell_{t+1}^{(q)} \in \arg\min_{\ell\geq 1} \frac{d^{\ell}}{\|\boldsymbol{\Gamma}_{\boldsymbol{U}_{\leq t}^{(q)},\ell}\|_{\mathrm{op}}}, \qquad \ell_{t+1}^{(s)} \in \arg\min_{\ell\geq 1} \frac{d^{\ell/2}}{\|\boldsymbol{\Gamma}_{\boldsymbol{U}_{\leq t}^{(s)},\ell}\|_{\mathsf{F}}}, \tag{39}$$

其中 $\boldsymbol{U}_{\leq t}^{(q)}$ and $\boldsymbol{U}_{\leq t}^{(s)}$ 表示在各个过程中恢复的子空间。根据构造，

$$\max_{t\in[T^{(q)}-1]} \frac{d^{\ell_t^{(q)}}}{\|\boldsymbol{\Gamma}_{\boldsymbol{U}_{\leq t}^{(q)},\ell_t^{(q)}}\|_{\mathrm{op}}} \lesssim \mathsf{qLeap}(\nu_d), \qquad \max_{t\in[T^{(s)}-1]} \frac{d^{\ell_t^{(s)}/2}}{\|\boldsymbol{\Gamma}_{\boldsymbol{U}_{\leq t}^{(s)},\ell_t}\|_{\mathsf{F}}} \lesssim \mathsf{sLeap}(\nu_d)$$

如果恢复 $\boldsymbol{U}_{t+1}^{(q)}$或 $\boldsymbol{U}_{(s)t+1}$的复杂度在每个阶段都达到统计查询和局部差分隐私下界，那么这些顺序过程就能达到相应的跃迁复杂度。

许多不同的次数序列都可能达到相同的总体复杂度。然而，如示例（第 1.2.3 节）所示，运行时间最优序列与样本最优序列可能导致截然不同的恢复轨迹。在这种情况下，必须选择优化运行时间还是样本量：我们的下界表明，不存在一个顺序过程能同时在两方面达到最优。此外，也可以选择介于这两个极端之间的次数序列，以实现中间权衡。

4 迭代式调和张量展开算法

本节介绍一系列顺序谱算法，用于利用球对称多指标模型的独立同分布样本恢复潜在子空间 $\boldsymbol{W}_*$。这些方法基于迭代应用调和张量展开算法（指定调和次数和核函数），并将其用于通过条件化已恢复的信号子空间而得到的约化球面 MIM。该过程中使用的调和次数序列可以任意选择，也可以遵循最优跃迁序列，根据样本跃迁（sLeap）或查询跃迁（qLeap）复杂度进行选择，详见第 3.3 节。

我们首先在第4.1节描述并分析一次调和张量展开。随后，我们将在第4.2 节展示，如何通过迭代这一过程，完整恢复潜在子空间 $\boldsymbol{W}_*$。

4.1 单步调和张量展开

令 $\{(\boldsymbol{y}_i,\ \boldsymbol{z}_i)\}_{i=1}^n$ 为从球对称多指标模型 $\mathbb{P}_{\nu_d}^{\boldsymbol{W}}$ 中抽取的独立同分布样本，其链接函数为ν_d，未知支撑集为 $\boldsymbol{W} \in \mathsf{Stf}_{\mathsf{s}}(\mathbb{R}^d)$。本节将介绍在固定调和次数 $\ell \geq 1$ 下执行一次调和张量展开的过程，并建立恢复

潜在信号方向的子空间 $\boldsymbol{U}_0 \subseteq \boldsymbol{W}$ 的保证。下文不失一般性地假设 $\|\boldsymbol{\xi}_{\emptyset,\,\ell}^{\boldsymbol{W}}\|_{L^2} > 0$；否则，调和次数 ℓ 上不存在信号，该过程无法识别任何潜在方向。为简化记号，在本节余下部分省略对$\boldsymbol{W}$的依赖，并记作 $\mathbb{P}_{\nu_d} := \mathbb{P}_{\nu_d}^{\boldsymbol{W}}$ 和 $\boldsymbol{\xi}_{\emptyset,\,\ell} := \boldsymbol{\xi}_{\emptyset\,\ell,\,}^{\boldsymbol{W}}$。

对于张量 $\boldsymbol{A} \in (\mathbb{R}^d)^{\otimes k}$，我们将其（$a$，$b$）展开定义为：对于满足 $a+b=k$ 的整数 a，$b \geq 0$，其对应的 $d^a \times d^b$ 矩阵 $\mathrm{Mat}_{a,\,b}(\boldsymbol{A})$ 的元素定义如下

$$(\mathrm{Mat}_{a,b}(\boldsymbol{A}))_{(i_1,\ldots,i_a),(j_1,\ldots,j_b)} = A_{i_1,\ldots,i_a,j_1,\ldots,j_b},$$

对所有多重指标（i_1，……，i_a）$\in [d]^a$和（j_1，……，j_b）$\in [d]^b$均成立。我们还引入正交归一标架 $\boldsymbol{U} \in \mathbb{R}^{d\times k}$ 与 $\boldsymbol{U}' \in \mathbb{R}^{d\times k}$之间的标准距离概念，定义如下

$$\mathrm{dist}(\boldsymbol{U},\boldsymbol{U}') := \|\boldsymbol{\Pi}_{\boldsymbol{U}} - \boldsymbol{\Pi}_{\boldsymbol{U}'}\|_{\mathrm{op}},$$

其中，$\boldsymbol{\Pi}_{\boldsymbol{U}} = \boldsymbol{U}\boldsymbol{U}^{\intercal}$表示投影到张成空间 $\mathrm{span}(\boldsymbol{U})$ 上的正交投影算子。最后，对于任意矩阵 $\boldsymbol{A} = [\boldsymbol{a}_1,\ldots,\boldsymbol{a}_q] \in \mathbb{R}^{p\times q}$，我们将其列空间记作 $\mathrm{span}(\boldsymbol{A}) := \mathrm{span}(\boldsymbol{a}_1$，……，$\boldsymbol{a}_q) \subseteq \mathbb{R}^p$。

4.1.1 预备知识

我们首先刻画通过一次调和张量展开算法步骤可以恢复的信号方向子空间 $\boldsymbol{U}_0 \subseteq \boldsymbol{W}$。

回顾一下，链接函数 ν_d 的第 ℓ 个调和系数定义为

$$\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}) := \mathbb{E}_{\mathbb{P}_{\nu_d}}[\mathcal{H}_\ell(\boldsymbol{z})|\boldsymbol{y}] \in \mathrm{TSym}_\ell(\mathbb{R}^d).$$

由于 $\mathbb{P}_{\nu_d}$ 在保持 $\boldsymbol{W}$ 不变的正交变换下保持不变，张量$\boldsymbol{\xi}_{\emptyset,\,\ell}(\boldsymbol{y})$ 本身也是 $\mathcal{O}_d^{\boldsymbol{W}}$-不变的，对所有 $\boldsymbol{y} \in \mathcal{Y}$均成立。因此，根据引理 20，存在一个张量值函数 $\boldsymbol{\zeta}_{\emptyset,\,\ell}(\boldsymbol{y}) \in \mathrm{Sym}_\ell(\mathbb{R}^{\mathsf{s}})$，使得

$$\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}) = \mathsf{P}_{\mathsf{tf}}(\boldsymbol{W}^{\otimes\ell}\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y})).$$

设 $D_{\mathsf{s},\,\ell} := \dim\big(\mathrm{Sym}_\ell(\mathbb{R}^{\mathsf{s}})\big)$，并定义

$$\boldsymbol{\Upsilon}_{\emptyset,\ell} := \mathbb{E}[\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y}) \otimes \boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y})] \in \mathrm{Sym}_\ell(\mathbb{R}^{\mathsf{s}}) \otimes \mathrm{Sym}_\ell(\mathbb{R}^{\mathsf{s}}). \tag{40}$$

该张量可自然地视为 $\mathrm{Sym}_\ell(\mathbb{R}^{\mathsf{s}})$ 上的自伴、半正定线性算子。根据谱定理，存在 $\mathrm{Sym}_\ell(\mathbb{R}^{\mathsf{s}})$ 的一组正交归一基 $\{\boldsymbol{V}_j\}_{j\in[D_{\mathsf{s},\ell}]}$ 以及非负特征值 $\mu_1 \geq \mu_2 \geq \cdots \geq 0$，使得

$$\boldsymbol{\Upsilon}_{\emptyset,\ell} = \sum_{j=1}^{D_{\mathsf{s},\ell}} \mu_j\, \boldsymbol{V}_j \otimes \boldsymbol{V}_j.$$

对于任意 $k \in [D_{\mathsf{s},\ell}]$，定义信号张量

$$\boldsymbol{Z}^{(k)} := \sum_{j=1}^{k} (\boldsymbol{W}^{\otimes\ell}\boldsymbol{V}_j) \otimes (\boldsymbol{W}^{\otimes\ell}\boldsymbol{V}_j) \in (\mathbb{R}^d)^{\otimes 2\ell}.$$

Let $\boldsymbol{U}^{(k)} \in \mathbb{R}^{d\times \mathsf{s}^{(k)}}$ 表示展开矩阵的左奇异向量矩阵

$$\mathrm{Mat}_{1,2\ell-1}(\boldsymbol{Z}^{(k)}) = \boldsymbol{W}\mathrm{Mat}_{1,2\ell-1}\left(\sum_{j=1}^{k} \boldsymbol{V}_j \otimes \boldsymbol{V}_j\right)[\boldsymbol{W}^{\otimes(2\ell-1)}]^{\intercal},$$

从而 $\boldsymbol{U}^{(k)} \subseteq \boldsymbol{W}$。这里，在轻微的记号滥用下，$\boldsymbol{W}^{\otimes(2\ell-1)}$ 表示 $d^{2\ell-1} \times \mathsf{s}^{2\ell-1}$矩阵，其元素由多重指标$(i_1, \ldots, i_{2\ell-1}) \in [d]^{2\ell-1}$与$(s_1, \ldots, s_{2\ell-1}) \in [\mathsf{s}]^{2\ell-1}$索引，由下式给出

$$(\boldsymbol{W}^{\otimes(2\ell-1)})_{(i_1,\ldots,i_{2\ell-1}),(s_1,\ldots,s_{2\ell-1})} = W_{i_1 s_1} W_{i_2 s_2} \ldots W_{i_{2\ell-1} s_{2\ell-1}}. \tag{41}$$

特别地，由于 $\boldsymbol{W}^{\intercal}\boldsymbol{W} = \mathbf{I}_{\mathsf{s}}$，我们有 $[\boldsymbol{W}^{\otimes(2\ell-1)}]^{\intercal}\boldsymbol{W}^{\otimes(2\ell-1)} = \mathbf{I}_{\mathsf{s}^\ell}$。

我们的结果表明，当 $\mu_{k+1} \ll \mu_k$ 时，样本量为 $n \asymp d^{\ell/2}/\mu_k$ 的单步估计量能够恢复由信号方向构成的子空间 $\boldsymbol{U}^{(k)}$。由于我们的目标是弱恢复，我们将 $k = \mathsf{r}$ 选为出现谱间隙时的最小指标，因此 $\mu_{\mathsf{r}} \asymp \mu_1 \asymp \|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2$ 且 $\mu_{\mathsf{r}+1} \ll \mu_{\mathsf{r}}$，相应的样本复杂度 $n \asymp d^{\ell/2}/\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2$ 与局部差分隐私下界相匹配。为简化表述，我们假设以下条件，该条件的成立不失一般性。

假设3 (谱间隙 $\gamma > 1$)。存在一个整数$\mathsf{r} \in [D_{\mathsf{s},\ell}]$ 和一个常数 $c_\mu \in (0, 1)$，使得算子 $\boldsymbol{\Upsilon}_{\emptyset,\ell}$ 的特征值 $\{\mu_j\}_{j\in[D_{\mathsf{s},\ell}]}$ 满足（约定 $\mu_{D_{\mathsf{s},\ell}+1} = 0$)

$$\mu_{\mathsf{r}} \geq c_\mu \mu_1 \quad \text{and} \quad \mu_{\mathsf{r}+1} \leq \frac{c_\mu \mu_1}{\gamma}.$$

注意，我们可以在不作此假设的情况下重述样本保证（定理 3），只需将（$\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2$ 和 γ（46）中的 μ_{r} 和 $\mu_{\mathsf{r}}/\mu_{\mathsf{r}+1}$ 分别替换为。对于假设3中的秩r的取值，我们记

$$\boldsymbol{Z}_0 := \boldsymbol{Z}^{(\mathsf{r})}, \qquad \boldsymbol{U}_0 := \boldsymbol{U}^{(\mathsf{r})}, \qquad \mathsf{s}_0 = \operatorname{rank}(\boldsymbol{U}_0). \tag{42}$$

在本节余下部分，我们将介绍并证明通过单步调和张量展开恢复 $\boldsymbol{U}_0$ 的保证。

4.1.2 调和张量展开算法

单步调和张量展开算法取决于满足 $a + b = \ell$ 的一组选定的形状参数（a, b）。这些参数决定如何将 ℓ 阶调和张量 $\mathcal{H}_{d,\ell}(\boldsymbol{z})$ 展开为矩阵。给定一个半正定核 $K\colon \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$，我们按如下方式构造经验矩阵 $\widehat{\boldsymbol{M}} \in \mathbb{R}^{d^a \times d^a}$：

- 若 ℓ 为偶数且 $a = b = \ell/2$,

$$\widehat{\boldsymbol{M}} = \frac{1}{n^2} \sum_{1\leq i,j\leq n} K(\boldsymbol{y}_i, \boldsymbol{y}_j) \mathrm{Mat}_{a,b}(\mathcal{H}_\ell(\boldsymbol{z}_i)) \mathrm{Mat}_{a,b}(\mathcal{H}_\ell(\boldsymbol{z}_j))^{\intercal}.$$

- 若 $\ell = 1$ 且（a, b）$=$（1, 0），或者 $\ell \geq 3$ 且 $1 \leq a < b$ 满足 $a + b = \ell$,

$$\widehat{\boldsymbol{M}} = \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} K(\boldsymbol{y}_i, \boldsymbol{y}_j) \mathrm{Mat}_{a,b}(\mathcal{H}_\ell(\boldsymbol{z}_i)) \mathrm{Mat}_{a,b}(\mathcal{H}_\ell(\boldsymbol{z}_j))^{\intercal}.$$

现在我们非正式地说明，如何从 $\widehat{\boldsymbol{U}}_0 \in \mathbb{R}^{d\times \mathsf{s}_0}$ 的前 $\widehat{}$ $\widehat{}$个特征向量估计 $\boldsymbol{M}$，使得 $\mathrm{dist}(\boldsymbol{U}_0, \boldsymbol{U}_0) = o(1)$，其中 $\boldsymbol{U}_0 \subseteq \boldsymbol{W}$是第 4.1.1 节中引入的信号子空间。对于合适的核 K（见下文的假设 4），可以证明

$$\widehat{\boldsymbol{M}} \approx \|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2 \mathrm{Mat}_{a,2\ell-1}(\boldsymbol{Z}_0) \mathrm{Mat}_{a,2\ell-1}(\boldsymbol{Z}_0)^{\intercal} + \boldsymbol{\Delta}, \tag{43}$$

其中误差项以高概率满足

$$\|\boldsymbol{\Delta}\|_{\mathrm{op}} \lesssim \frac{d^{\ell/2}}{n} + \frac{d^{\ell/4}\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}}{\sqrt{n}}.$$

算法 1： 使用分解（a，b）的单步谱波张量展开 $a+b=\ell$

Function `TensorUnfoldOneStep`$(\{(\boldsymbol{y}_i,\boldsymbol{z}_i)\}_{i=1}^n$, *degree* ℓ, *kernel* K, *rank* t, *rank* $\mathsf{s}_0)$

`// Compute the empirical matrix` $\widehat{\boldsymbol{M}}\in\mathbb{R}^{d^a\times d^a}$

if $a=b$ **then**

$\widehat{\boldsymbol{M}}\leftarrow\frac{1}{n^2}\sum_{1\le i,j\le n}K(\boldsymbol{y}_i,\boldsymbol{y}_j)\mathrm{Mat}_{a,a}(\mathcal{H}_\ell(\boldsymbol{z}_i))\mathrm{Mat}_{a,a}(\mathcal{H}_\ell(\boldsymbol{z}_j))^\mathsf{T}$

else

$\widehat{\boldsymbol{M}}\leftarrow\frac{1}{n(n-1)}\sum_{1\le i\ne j\le n}K(\boldsymbol{y}_i,\boldsymbol{y}_j)\mathrm{Mat}_{a,b}(\mathcal{H}_\ell(\boldsymbol{z}_i))\mathrm{Mat}_{a,b}(\mathcal{H}_\ell(\boldsymbol{z}_j))^\mathsf{T}$

`// Extract the top t eigenvectors`

$\widehat{\boldsymbol{V}}=[\widehat{\boldsymbol{v}}_1,\dots,\widehat{\boldsymbol{v}}_\mathsf{t}]\leftarrow$ top t eigenvectors of $\widehat{\boldsymbol{M}}$; `//` $\widehat{\boldsymbol{V}}\in\mathbb{R}^{d^a\times\mathsf{t}}$

`// Contract to a` d`-dimensional subspace`

$\widehat{\boldsymbol{U}}_0\leftarrow$ top s_0 eigenvectors of $\sum_{s=1}^\mathsf{t}\mathrm{Mat}_{1,a-1}(\widehat{\boldsymbol{v}}_s)\mathrm{Mat}_{1,a-1}(\widehat{\boldsymbol{v}}_s)^\mathsf{T}$; `//` $\widehat{\boldsymbol{U}}_0\in\mathbb{R}^{d\times\mathsf{s}_0}$

return $\widehat{\boldsymbol{U}}_0$

令 $\mathsf{t}=\mathrm{rank}(\mathrm{Mat}_{a,2\ell-a}(\boldsymbol{Z}_0))$，并令 $\boldsymbol{V}=[\boldsymbol{v}_1,\dots,\boldsymbol{v}_\mathsf{t}]\in\mathbb{R}^{d^a\times\mathsf{t}}$ 表示 Mat_a 的前 t 个左特征向量$_{a,2\ell-a}(\boldsymbol{Z}_0)$。根据 (43)，当 $n\gtrsim d^{\ell/2}/\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2$ 时，计算

$$\widehat{\boldsymbol{V}}=[\widehat{\boldsymbol{v}}_1,\dots,\widehat{\boldsymbol{v}}_\mathsf{t}]\in\mathbb{R}^{d^a\times\mathsf{t}}\ \leftarrow\ \text{top }\mathsf{t}\text{ eigenvectors of }\widehat{\boldsymbol{M}}$$

得到 $\mathrm{span}(\widehat{\boldsymbol{V}})\approx\mathrm{span}(\boldsymbol{V})$。为提取 $\mathbb{R}^d$ 中的子空间，我们进一步对这些特征向量进行收缩并计算

$$\widehat{\boldsymbol{U}}_0\leftarrow\text{top }\mathsf{s}_0\text{ eigenvectors of }\sum_{s=1}^\mathsf{t}\mathrm{Mat}_{1,a-1}(\widehat{\boldsymbol{v}}_s)\mathrm{Mat}_{1,a-1}(\widehat{\boldsymbol{v}}_s)^\mathsf{T}.$$

所得子空间 $\mathrm{span}(\widehat{\boldsymbol{U}}_0)$ 接近 $\sum_{s=1}^\mathsf{t}\mathrm{Mat}_{1,\,a-1}(\boldsymbol{v}_s)\mathrm{Mat}_{1,\,a-1}(\boldsymbol{v}_s)^\mathsf{T}$ 的张成空间，而该空间恰好与 $\mathrm{span}(\boldsymbol{U}_0)$ 一致。完整过程总结于算法 1。

算法的样本复杂度。我们假设核满足以下条件。

假设 4 (可容许核)。半正定核 $K:\ \mathcal{Y}\times\mathcal{Y}\to\mathbb{R}$ 若满足以下条件，则称其为可容许核*s:*

(K1) (有界性) 存在常数 $B_K>0$，使得 $\|K\|_\infty\le B_K$.

(K2) (谱 信号方向下界) 存在常数 $c_K>0$ 使得 $t,$

$$\mathbb{E}\left[K(\boldsymbol{y},\boldsymbol{y}')\langle\boldsymbol{V},\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y})\rangle_\mathsf{F}\langle\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y}'),\boldsymbol{V}\rangle_\mathsf{F}\right]\ge c_K\mathbb{E}\left[\langle\boldsymbol{V},\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y})\rangle_\mathsf{F}\langle\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y}),\boldsymbol{V}\rangle_\mathsf{F}\right],\qquad(44)$$

对所有 $\boldsymbol{V}\in\mathrm{TSym}_\ell(\mathbb{R}^s)$，其中 $(\boldsymbol{y},\boldsymbol{y}')\sim\nu_d^Y\otimes\nu_d^Y$。

(K3) (有限秩) 存在 $\mathsf{m}\in\mathbb{N}$，使得 K 的秩为 m，也就是说，存在 m 个函数$\mathcal{T}_s:\mathcal{Y}\to\mathbb{R}$，使得

$$K(\boldsymbol{y},\boldsymbol{y}')=\sum_{s=1}^\mathsf{m}\mathcal{T}_s(\boldsymbol{y})\mathcal{T}_s(\boldsymbol{y}').$$

假设4.(K1) 用于控制经验矩阵的集中性 $\boldsymbol{M}$ 及其期望。假设4.(K2) 确保期望 $\mathbb{E}[\boldsymbol{M}]$ 在信号方向上是正定的，从而使算法能够恢复信号子空间。最后，假设4.(K3)使运行时间能够与统计查询框架的预测相匹配。

备注 4.1 (预言机核). 理想情况下，我们希望使用如下定义的预言机核:

$$K(\boldsymbol{y},\boldsymbol{y}') := \boldsymbol{\lambda}(\boldsymbol{y})^{\mathsf{T}}\mathbb{E}[\boldsymbol{\lambda}(\boldsymbol{y})\boldsymbol{\lambda}(\boldsymbol{y}')^{\mathsf{T}}]^{\dagger}\boldsymbol{\lambda}(\boldsymbol{y}'), \qquad \text{where} \quad \boldsymbol{\lambda}(\boldsymbol{y}) := \mathrm{Mat}_{\ell,0}(\boldsymbol{\zeta}_{\emptyset,\ell}) \in \mathbb{R}^{\mathsf{s}^{\ell}} \tag{45}$$

采用这一选择后，式（44）取等号，其中 $c_K = 1$，而 K 的秩至多为 s^{ℓ}。然而，该核可能不满足有界性条件（假设 4.(K1)），这是我们分析中限制最严格的要求。我们可以将这一条件放宽为对 K满足适当的矩界，代价是样本复杂度中增加一个 polylog(d) 因子。此外，在 ν_d 未知（或仅部分已知）的情况下，考虑满足$\mathsf{m} = \infty$的稠密核可能更合适。稍作修改后，我们的分析也可扩展到这种情况（见附录 B.1.2）。我们还在注释 ν_d 中进一步讨论了 4.2 未知的情况。

在假设 4下，我们得到以下保证：利用从 $\mathbb{P}_{\nu_d}$ 中独立同分布抽取的样本，可以恢复信号子空间 $\boldsymbol{U}_0$。

定理 *3*（单步谐波张量展开）。设 $(\nu_d, \boldsymbol{W})$ 为球面多指标模型，且 $\ell \geq 1$ 为满足$\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2} > 0$的整数。在假设*3*（间隙为 $\gamma > 1$ ）和假设 *4* 下，存在常数 $C, C', c, c' > 0$，它们仅依赖于 ℓ、s 以及这些假设中的常数，使得以下结论成立。设 $\boldsymbol{U}_0 \subseteq \boldsymbol{W}$ 为式（*42*）中定义的信号子空间，且令 $\widehat{\boldsymbol{U}}_0 \in \mathit{Stf}_{\mathsf{s}_0}(\mathbb{R}^d)$ 为算法 *1* 的输出。那么，对任意 $C' \leq \gamma \leq \sqrt{d}$，以及 $n \leq \exp(d^{c'})$，

$$\mathrm{dist}(\widehat{\boldsymbol{U}}_0, \boldsymbol{U}_0) \leq C\left[\sqrt{\frac{d^{\ell/2\vee 1}}{n\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2}} + \frac{1}{\sqrt{\gamma}}\right], \tag{46}$$

以至少 $1 - \exp(-d^c)$ 的概率。

定理3 的证明见附录B.1 ，对应非对称展开情形（$a \neq b$）；对称情形（$a = b$）的证明见附录 γ。取B.2 充分大，则界（46）意味着，可以通过样本量 $\boldsymbol{U}_0$ 恢复

$$n \asymp \frac{d^{\ell/2\vee 1}}{\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2}.$$

对于 $\ell \geq 2$，无论选择何种展开 $1 \leq a \leq b$，这都与局部差分隐私下界（35）匹配。对于 $\ell = 1$，与定理1(a)中的检测下界相比，该界差了一个因子 $\sqrt{d}$。然而，正如注3.1所述，这反映了此情形中固有的检测-恢复差距。当 $\|\boldsymbol{\xi}_{\emptyset,1}\|_{L^2}^2 \asymp 1$ 时，样本复杂度 $n \asymp d$ 实际上是信息论最优的。当$\|\boldsymbol{\xi}_{\emptyset,1}\|_{L^2}^2 \ll 1$时，沿用 [JKMS25], 的思路，可以证明算法能够恢复一组方向，其重叠度为 $\widetilde{\Theta}_d(d^{-1/4})$; 在附加假设下，利用高阶谐波可将其提升至 $\Theta_d(1)$，从而匹配检测下界。若不作此类假设，要确定 $\ell = 1$ 时恢复的最优样本复杂度，就需要建立估计的局部差分隐私下界（例如，沿着 [SW22,CGGV25] 的思路）。我们将此问题留待未来研究。最后请注意，即使假设3在 $\gamma = \infty$ 下成立，定理3中的保证（46）也只能得到 $\mathrm{dist}(\widehat{\boldsymbol{U}}_0, \boldsymbol{U}_0) \lesssim d^{-1/4}$，当 $n \to \infty$ 时。要达到任意精度，需要采用不同的算法进行第二次细化。

算法的运行时间。 算法的主要计算成本 1 在于计算 $\widehat{\ }$经验矩阵的主特征向量 $\boldsymbol{M}$。这通过子空间幂法

迭代完成，整体运行时间取决于向量与 $\boldsymbol{M}$相乘的计算成本。在有限秩假设（假设 4。(K3)）下，矩阵 $\boldsymbol{M}$ 可分解为

$$\widehat{\boldsymbol{M}} = \left(1 + \frac{\delta_{a\neq b}}{n-1}\right) \sum_{r=1}^{\mathsf{m}} \left[\boldsymbol{S}_r \boldsymbol{S}_r^{\mathsf{T}} - \delta_{a\neq b} \boldsymbol{D}_r\right], \tag{47}$$

其中

$$\boldsymbol{S}_r := \frac{1}{n} \sum_{i\in[n]} \mathcal{T}_r(\boldsymbol{y}_i) \mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_i)), \qquad \boldsymbol{D}_r := \frac{1}{n^2} \sum_{i\in[n]} \mathcal{T}_r(\boldsymbol{y}_i)^2 \mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_i)) \mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_i))^{\mathsf{T}}.$$

因此，将向量与 $\boldsymbol{M}$ 相乘可归结为计算涉及 $\boldsymbol{S}_r$ 和 $\boldsymbol{D}_r$ 的矩阵-向量乘积，每个乘积需要对 n 项求和，而非 n^2。此外，利用调和张量的结构，可以在 $O(d^a + d^b)$ 时间内计算这些矩阵 - 向量乘积，而无需显式构造完整矩阵 $\mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}))$；否则需要 $O(d^{a+b})$ 次运算。由此可得以下运行时间保证。

命题 1 (调和张量展开的运行时间)。 存在一个常数 $C > 0$，其仅取决于 ℓ 和 s，使得对于所有 $\varepsilon > 0$，均可在精度 $\boldsymbol{U}_0$ 内近似算法 1的输出 $\varepsilon > 0$ （相对于 dist），所需时间为

$$T \leq C \frac{\mathsf{mt}}{g} n(T_K + d^a + d^b) \log(d/\varepsilon),$$

其中 T_K 表示计算 $\{\mathcal{T}_r\}_{r\in[\mathsf{m}]}$ 的代价，且 $g = 1 - \lambda_{\mathsf{t}+1}/\lambda_{\mathsf{t}}$ 满足 λ_j 是 j 阶特征值，属于 $\boldsymbol{M}$。特别地，在定理 3的条件下，$g \geq 1 - c/\gamma$ 以高概率成立。

运行时间分析的更多细节见附录 B.4。虽然对于 $a \leq b$，样本复杂度与展开形状无关，但运行时间取决于这一选择：展开形状越接近正方形，计算成本越低。因此，我们取 $a = \lfloor \ell/2 \rfloor$和 $b = \lceil \ell/2 \rceil$，这对应于偶数 $a = b = \ell/2$ 时的 ℓ，以及奇数 a时的 $(b) = ((\ell-1)/2, (\ell+1)/2)$。$\ell$

结合 定理 3与命题1，信号子空间 $\boldsymbol{U}_0$ 可以在

$$\mathsf{T} \asymp \begin{cases} d^2 \log(d)/\|\boldsymbol{\xi}_{\emptyset,1}\|_{L^2}^2 & \text{if } \ell = 1; \\ d^\ell \log(d)/\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2 & \text{if } \ell \text{ is even}; \\ d^{\ell+\frac{1}{2}} \log(d)/\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2 & \text{if } \ell \text{ is odd}. \end{cases}$$

对于偶数 ℓ，这与 SQ 运行时间下界相符（相差一个对数因子），但对于奇数$\sqrt{d}$，运行时间则相差[9] $\ell \geq 3$倍。能否消除这一差距仍是一个开放问题。对于单指标模型这一特殊情形（$\mathsf{s} = 1$），[JKMS 25] 证明了，对于所有 $\ell \geq 3$，在线随机梯度下降可实现运行时间 $d^\ell/\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2$ 不含对数因子），但样本复杂度更高，为 $d^{\ell-1}/\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2$。我们将实现奇数 ℓ情形下最优 SQ 运行时间的问题留待未来研究。

注4.2 (未知 ν_d)。 算法1 依赖于对 ν_d 的了解，这体现在核 K的选择以及秩t 和s_0的选择上。当 ν_d 仅部分已知时，可以改为选择一个稠密且有界的核，使其适用于一类合适的链接函数，从而满足假设 4.(K2)

[9]对于 $\ell = 1$，其表现差了 d 倍。在 $\|\boldsymbol{\xi}_{\emptyset,1}\|_{L^2} \asymp d$ 的情况下，这是最优的：它对应于在 $\mathbb{R}^d$ 中读取 $\Theta_d(d)$ 个样本所需的运行时间，而这在信息论意义上是必需的。

成立。一个自然的例子是正核，例如高斯核，这在 [DLB25]中已有考虑。然而，我们注意到， [DLB25]的设定与我们的不同：其中标签不依赖于环境维数，而在本文中，对 d 的依赖带来了额外挑战。我们在此不再进一步探讨这一方向。此类稠密核通常不具有有限秩表示，因此无法进行高效分解（47），且不满足我们对可容许性的定义（见假设 4.(K3)）。尽管如此，可以采用随机特征近似，以恢复计算效率，其表现几乎与我们的统计查询下界相匹配。最后，当 t和 s_0 未知时，可以自适应地选取这些量，例如仅保留与主体谱充分分离的特征值。

4.2 多步过程

现在我们说明，如何将单步调和张量展开算法迭代应用于一系列约化多指标模型，从而恢复整个支撑集 $\boldsymbol{W}$。

我们首先固定一个调和次数序列 $\ell_1,\ \ell_2,\ \ell_3,\ \ldots \geq 1$，并描述与这一选择相关的算法。随后，我们讨论如何选择次数序列，以实现最优的样本复杂度和运行时间复杂度（在局部差分隐私和统计查询框架下）。该算法依次估计一组子空间 $\{\widehat{\boldsymbol{U}}_t\}_{t=1}^{T}$。记 $\mathsf{s}_0 = \mathsf{s}_{\leq 0} = 0$ 和 $d_0 := d$，并定义 for $t \geq 1$

$$\mathsf{s}_t := \operatorname{rank}(\widehat{\boldsymbol{U}}_t), \qquad \mathsf{s}_{\leq t} := \mathsf{s}_1 + \ldots + \mathsf{s}_t, \qquad d_t := d - \mathsf{s}_{\leq t-1},$$

每个子空间 $\widehat{\boldsymbol{U}}_t \in \mathrm{Stf}_{\mathsf{s}_t}(\mathbb{R}^{d_t})$ 都被视为位于先前恢复的方向 $\widehat{\ }\boldsymbol{U}_{\leq t-1}$ 的正交补中，其中这些方向由下式给出

$$\widehat{\boldsymbol{U}}_{\leq t-1} := \widehat{\boldsymbol{U}}_1 \oplus \ldots \oplus \widehat{\boldsymbol{U}}_{t-1} \in \mathrm{Stf}_{\mathsf{s}_{\leq t-1}}(\mathbb{R}^d).$$

为介绍该算法，我们首先定义一个理想化的总体恢复序列 $\{\boldsymbol{U}_t\}_{t=1}^{T}$，然后说明如何利用调和张量展开从数据中近似恢复这一序列。

总体恢复序列。 固定一个谱隙参数 $\gamma > 0$，并设 $\boldsymbol{U}_0 := \emptyset$。我们递归地构造序列 $\{\boldsymbol{U}_t\}_{t=1}^{T}$ 如下。假设 $\{\boldsymbol{U}_k\}_{k=0}^{t}$ 已定义，其中 $\mathsf{s}_k :=$秩$(\boldsymbol{U}_k)$。定义

$$\boldsymbol{U}_{\leq t} := \boldsymbol{U}_0 \oplus \ldots \oplus \boldsymbol{U}_t \in \mathrm{Stf}_{\mathsf{s}_{\leq t}}(\mathbb{R}^d), \qquad \mathsf{s}_{\leq t} := \mathsf{s}_0 + \ldots + \mathsf{s}_t, \qquad d_{t+1} := d - \mathsf{s}_{\leq t}.$$

我们考虑通过对 $\boldsymbol{U}_{\leq t}^{\mathsf{T}}\boldsymbol{z}$ 条件化得到的约化多指标模型，如 (6) 中所述。令$\boldsymbol{U}_{\leq t,\perp} \in \mathbb{R}^{d\times d_{t+1}}$ 为 $\boldsymbol{U}_{\leq t}$ 的正交补的一个正交归一基，并将数据 $(\boldsymbol{y}, \boldsymbol{z})$ 分解为

$$\boldsymbol{r}_t := \boldsymbol{U}_{\leq t}^{\mathsf{T}}\boldsymbol{z} \in \mathbb{R}^{\mathsf{s}_{\leq t}}, \qquad \boldsymbol{z}_t := \frac{\boldsymbol{U}_{\leq t,\perp}^{\mathsf{T}}\boldsymbol{z}}{\|\boldsymbol{U}_{\leq t,\perp}^{\mathsf{T}}\boldsymbol{z}\|_2} \in \mathbb{S}^{d_{t+1}-1}, \qquad \boldsymbol{y}_t := (\boldsymbol{y}, \boldsymbol{r}_t) \in \mathcal{Y}_t := \mathcal{Y} \times \mathbb{R}^{\mathsf{s}_{\leq t}}.$$

记 $\nu_{d,t} := \nu_{d,\boldsymbol{U}_{\leq t}}$ 为与该分解对应的约化 MIM 的链接函数。在给定 $\boldsymbol{z}_t$ 的条件下，响应满足

$$\boldsymbol{y}_t|\boldsymbol{z}_t \sim \nu_{d,t}(\mathrm{d}\boldsymbol{y}_t|\boldsymbol{W}^{\mathsf{T}}\boldsymbol{U}_{\leq t,\perp}\boldsymbol{z}_t) = \nu_d(\mathrm{d}\boldsymbol{y}|\boldsymbol{W}^{\mathsf{T}}\boldsymbol{z})\tilde{\tau}_{d,\mathsf{s}_{\leq t}}(\mathrm{d}\boldsymbol{r}_t).$$

我们将次数为ℓ_{t+1} 的调和张量展开过程应用于该约化多重指标模型，它有$\mathsf{s} - \mathsf{s}_{\leq t}$ 个剩余指标。令r_{t+1} 表示满足假设3 且间隙为 γ（特别是，

）$c_\mu > \gamma^{-\mathsf{r}_{t+1}}$。接着定义 $\boldsymbol{U}_{t+1} \in \mathrm{Stf}^{\mathsf{s}_{t+1}}(\mathbb{R}^{d_{t+1}})$ 为根据（42）构造的相应单步过程的信号子空间。我们将关联的信号张量记为 $\boldsymbol{Z}_{t+1}$，并写作 $\mathsf{t}_{t+1} := \mathsf{rank}(\mathrm{Mat}_{a,\ 2\ell_{t+1}-a}(\boldsymbol{Z}_{t+1})$ 。约化多重指标模型的第 ℓ_{t+1} 个调和系数为

$$\boldsymbol{\xi}_t(\boldsymbol{y}_t) := \boldsymbol{\xi}_{\boldsymbol{U}_{\le t},\ell_{t+1}}(\boldsymbol{y}) = \mathbb{E}[\mathcal{H}_{d_{t+1},\ell_{t+1}}(\boldsymbol{z}_t)|\boldsymbol{y}_t] \in \mathrm{TSym}_{\mathsf{s}-\mathsf{s}_{\le t}}(\mathbb{R}^{d_{t+1}}).$$

不失一般性，假设对于所有 $t = 0, \dots, T-1$，$\|\boldsymbol{\xi}_t\|_{L^2} > 0$，并令 T 满足 $\boldsymbol{U}_{\le T} = \boldsymbol{W}$（特别地，$T \le \mathsf{s}$）。最后，假设存在一个核序列$\{K_t\}_{t=1}^{T}$，使得每个 K_t： $\mathcal{Y}_{t-1} \times \mathcal{Y}_{t-1} \to \mathbb{R}$ s 都满足假设 4，该假设针对的是这对 $(\nu_{d,t-1}, \ell_t)$。

经验恢复序列。 现在我们描述经验多步过程：该过程根据总体恢复序列选择参数，并在每一步使用新的样本，迭代地应用调和张量展开。

初始化 $\widehat{\boldsymbol{U}}_0 := \emptyset$。对于 $t = 0, \dots, T-1$，假设我们已获得估计 $\{\widehat{\boldsymbol{U}}_k\}_{k=0}^{t}$，其秩 $\mathsf{rank}(\boldsymbol{U}_k)$ 为 $= \mathsf{s}_k$。接着，我们对约化多指标模型 $(\hat{\boldsymbol{y}}_t, \hat{\boldsymbol{z}}_t) \sim \hat{\nu}_{d,t}$ 在给定 $\boldsymbol{U}_{\le t}$ 的条件下应用算法 1：

$$\hat{\nu}_{d,t} := \nu_{d,\widehat{\boldsymbol{U}}_{\le t}}, \qquad \hat{\boldsymbol{y}}_t := (\boldsymbol{y}, \boldsymbol{r}_{\widehat{\boldsymbol{U}}_{\le t}}), \qquad \hat{\boldsymbol{z}}_t := \boldsymbol{z}_{\widehat{\boldsymbol{U}}_{\le t}}.$$

算法在调和次数 ℓ_{t+1} 下运行，并采用最优展开形状：若$\ell_{t+1} = 1$，则为 $(a, b) = (1,0)$，否则为 $(a, b) = (\lfloor \ell_{t+1}/2 \rfloor, \lceil \ell_{t+1}/2 \rceil)$。我们使用目标秩 t_{t+1} 和 s_{t+1}，以及核

$$\overline{K}_{t+1}((\boldsymbol{y}, \boldsymbol{r}), (\boldsymbol{y}', \boldsymbol{r}')) = \int_{\mathcal{O}_{\mathsf{s}_{\le t}}} K_{t+1}((\boldsymbol{y}, g \cdot \boldsymbol{r}), (\boldsymbol{y}', g \cdot \boldsymbol{r}'))\pi_{\mathsf{s}_{\le t}}(\mathrm{d}g). \tag{48}$$

使用该对称化核，是因为该算法仅恢复 $\boldsymbol{U}_{\le t+1}$的张成空间，而不一定是某个特定的基。更确切地说，该算法返回一个子空间 $\widetilde{\boldsymbol{U}}_{t+1} \subseteq \mathbb{R}^{d_{t+1}}$，该子空间位于$\boldsymbol{U}_{\le t,\ \perp}^{\top}\boldsymbol{W}_*$的像中。我们将该估计提升回原来的 d维空间，方法是取 $\widehat{\boldsymbol{U}}_{t+1}$ 为$\boldsymbol{U}_{\le t,\ \perp}\widetilde{\boldsymbol{U}}_{t+1} \subseteq \mathbb{R}^d$的像的任意正交归一基。更多讨论见附录 B.3。

完整的多步过程总结于算法 2中。

恢复保证。 为了证明上述过程确实能够恢复完整的信号子空间 $\boldsymbol{W}$，我们希望在每次迭代中应用定理 3 。然而，在第一步之后无法直接这样做。事实上，对于 $t \ge 1$，我们只有 $\boldsymbol{U}_{\le t}$ 的近似估计，而经验约化模型 $\hat{\nu}_{d,\ t} := \nu_{d,\ \widehat{\boldsymbol{U}}_{\le t}}$ 不同于总体模型 $\nu_{d,\ t} := \nu_{d,\ \boldsymbol{U}_{\le t}}$。特别是，由于 $\boldsymbol{U}_{\le t}$ 并非 $\boldsymbol{W}$ 的一个确切子空间，模型 $\hat{\nu}_{d,\ t}$ 几乎必然仍然依赖于 s 个指标。

为了控制迭代过程中产生的误差传播，我们对约化链接函数族 $\{\nu_{d,t}\}_{t=0}^{T-1}$施加稳定性条件。非正式地说，该条件要求约化模型的调和系数相对于条件子空间的扰动连续变化。

假设 5（约化链接函数的稳定性）。 存在一个连续模 $\varphi : [0,1] \to \mathbb{R}_{\ge 0}$，其中 $\varphi(\varepsilon) \to 0$ 当 $\varepsilon \to 0$ 时，使得以下条件成立。

(a) 对所有 $t \in [T]$和 $\boldsymbol{U}' \in \mathrm{Stf}^{\mathsf{s}_{\le t-1}}(\mathbb{R}^d)$，令 $\boldsymbol{U}'_\perp$表示其正交补，

$$\|\boldsymbol{U}_{\le t-1,\perp}^{\otimes \ell_t}\boldsymbol{\xi}_{\boldsymbol{U}_{\le t-1},\ell_t} - (\boldsymbol{U}'_\perp)^{\otimes \ell_t}\boldsymbol{\xi}_{\boldsymbol{U}',\ell_t}\|_{L^2} \le (\|\boldsymbol{\xi}_{\boldsymbol{U}_{\le t-1},\ell_t}\|_{L^2} + \|\boldsymbol{\xi}_{\boldsymbol{U}',\ell_t}\|_{L^2})\varphi\big(\mathrm{dist}(\boldsymbol{U}_{\le t-1}, \boldsymbol{U}')\big).$$

其中 $\boldsymbol{U}_{\le t,\ \perp}$ 是总体恢复序列中 $\boldsymbol{U}_{\le t}$ 的正交补。

算法 2： 球面 MIM 的多步谐波张量展开

1 函数 MultiStepTensorUnfolding($\{\{(\boldsymbol{y}_i^{(t)}, \boldsymbol{z}_i^{(t)})\}_{i=1}^n\}_{t=0}^{T-1}$**，次数** $\{\ell_t\}_{t=1}^T$**，核函数** $\{K_t\}_{t=1}^T$**，秩** $\{\mathbf{t}_t\}_{t=1}^T$**，秩** $\{\mathbf{s}_t\}_{t=1}^T$**) // 初始化已恢复子空间** $\widehat{\boldsymbol{U}}_0 \leftarrow \emptyset$ **// 遍历各步骤 3 对于** $t \leftarrow 0$ **至** $T-1$ **执行 // 构造残差子空间4** $\widehat{\boldsymbol{U}}_{\le t} \leftarrow \bigoplus_{j=0}^{t} \widehat{\boldsymbol{U}}_j$ **// 在残差空间中对一批样本** t**进行分解5** $\{(\hat{\boldsymbol{y}}_{t,i}, \hat{\boldsymbol{z}}_{t,i})\}_{i=1}^n \leftarrow$ **在给定** $\widehat{\boldsymbol{U}}_{\le t}$**的条件下对** $\{(\boldsymbol{y}_i^{(t)}, \boldsymbol{z}_i^{(t)})\}_{i=1}^n$ **进行分解// 使用** $(a,b) = (1,0)$ **或** $(\lfloor \frac{\ell_{t+1}}{2} \rfloor, \lceil \frac{\ell_{t+1}}{2} \rceil)$**执行一步张量展开6** $\boldsymbol{U}_{t+1} \leftarrow$ **TensorUnfoldOneStep(**$\{(\hat{\boldsymbol{y}}_{t,i}, \hat{\boldsymbol{z}}_{t,i})\}_{i=1}^n, \ell_{t+1}, K_{t+1}, \mathbf{t}_{t+1}, \mathbf{s}_{t+1}$**) 7 返回** $\widehat{\boldsymbol{W}} = \bigoplus_{j=1}^{T} \widehat{\boldsymbol{U}}_j$

(b) 对所有 $t = 0、\ldots、T-1$ 和 $\boldsymbol{U}' \in \mathrm{Stf}^{\mathbf{s}}_{\le t}(\mathbb{R}^d)$，

$$\mathbb{E}\left[\left(K_{t+1}\left(\boldsymbol{y}_t, \boldsymbol{y}_t'\right) - K_{t+1}\left(\hat{\boldsymbol{y}}_t, \hat{\boldsymbol{y}}_t'\right)\right)^2\right]^{1/2} \le \varphi(\|\boldsymbol{U}_{\le t} - \boldsymbol{U}'\|_{\mathsf{F}}),$$

其中 $\boldsymbol{y}_t = (\boldsymbol{y}, \boldsymbol{U}_{\le t}^{\mathsf{T}}\boldsymbol{z})$、$\hat{\boldsymbol{y}}_t = (\boldsymbol{y}、(\boldsymbol{U}')^{\mathsf{T}}\boldsymbol{z})$ 和 $(\boldsymbol{y}_t'、\hat{\boldsymbol{y}}_t')$ 表示相互独立的副本。

在假设 5.(a) 下，式（45）中定义的预言机核本身对于条件子空间的扰动是稳定的。然而，由于算法不会直接使用该预言机核，我们还对核序列 $\{K_t\}_{t=1}^T$ (假设 5.(b)) 施加稳定性条件。尽管假设 5 有些严格，但高斯多指标模型等重要示例满足该假设；将输入向量归一化为单位范数后得到的高斯 MIMs 模型也满足该假设。

在这些条件下，我们得到以下多步过程的恢复保证。

定理*4*（多步谐波张量展开）。设 $(\nu_d, \boldsymbol{W})$ 是一个球面多指标模型，且 $(\ell_t)_{t\ge1}$ 是一个调和次数序列。设 $\{\boldsymbol{U}_t\}_{t=1}^T$ 是相应的总体恢复序列，其谱间隙为 $\gamma > 1$，并设 $\widehat{\boldsymbol{W}}$ 是算法 *2* 的输出。假设核序列 $\{K_t\}_{t=1}^T$ 满足假设 *4*，且假设 *5* 成立。则存在常数C、C'、c、$c' > 0$，它们仅取决于 s、次数序列以及这些假设中的常数，使得对于任意 $C' \le \gamma \le \sqrt{d}$，且 $n \le \exp(d^{c'})$，

$$\mathrm{dist}(\widehat{\boldsymbol{W}}, \boldsymbol{W}) \le C\left[\max_{t\in[T]}\left(\frac{d^{\ell_t/2\vee1}}{n\|\boldsymbol{\xi}_{t-1}\|_{L^2}^2}\right)^{1/C} + \frac{1}{\gamma^{1/C}}\right],$$

其概率至少为 $1 - \exp(-d^c)$。

定理 4 的证明见附录B.3。该定理表明，取 γ足够大时，使用次数序列为 $\boldsymbol{W}$ 的迭代张量展开来恢复信号子空间，其样本复杂度为 $\{\ell_t\}_{t\ge1}$ 是

$$n \asymp \max_{t\in[T]} \frac{d^{\ell_t/2\vee1}}{\|\boldsymbol{\xi}_{t-1}\|_{L^2}^2}.$$

关于运行时间复杂度，虽然 K_t 被假设为有限秩，但对称化核（48）不一定保留这种低秩结构。在附录 B.3中，我们提出了一种构造有限秩近似的方法。利用该近似，得到的总运行时间复杂度为

$$\mathsf{T} \asymp \max_{t\in[T]} \frac{d^{\ell_t \vee (3/2)+\frac{1}{2}\delta_{\ell_t\equiv 1[2]}}}{\|\boldsymbol{\xi}_{t-1}\|_{L^2}^2}\log(d).$$

最优多步过程。 原则上，算法 2可以采用任意调和次数序列运行 $\{\ell_t\}_{t=1}^{T}$。根据定理 2及其后续讨论所确立的下界，可以选择调和次数，使其在局部差分隐私和统计查询框架下将样本复杂度或运行时间降至最低，也可以在这两个目标之间进行权衡。

具体而言，考虑样本最优次数序列和运行时间最优次数序列 $\{\ell_t^{(s)}\}_{t=1}^{T_s}$和 $\{\ell_t^{(q)}\}_{t=1}^{T_q}$，它们定义于式（39），以及相应的总体恢复序列$\{\boldsymbol{U}_{(t}^{\ s}\}_{t=1}^{T_s}$）和 $\{\boldsymbol{U}_t^{(q)}\}_{t=1}^{T_q}$。根据定理 4，采用这两种次数序列的多步调和张量展开算法可分别达到样本复杂度和运行时间复杂度

$$n \asymp \max_{t\in[T_s]} \inf_{\ell\ge 1} \frac{d^{\ell/2\vee 1}}{\|\boldsymbol{\xi}_{\boldsymbol{U}_{\le t-1}^{(s)},\ell}\|_{L^2}^2}, \qquad \mathsf{T} \asymp \max_{t\in[T_q]} \inf_{\ell\ge 1} \frac{d^{\ell\vee(3/2)+\frac{1}{2}\delta_\ell}}{\|\boldsymbol{\xi}_{\boldsymbol{U}_{\le t-1}^{(q)},\ell}\|_{L^2}^2}\log(d),$$

其中，若 $\delta_\ell = 1$ 为 ℓ 奇数则取 1，否则取 0。这与定理 2中的样本跃迁复杂度和查询跃迁复杂度相符，至多相差一些 $\widetilde{O}(\sqrt{d})$ 因子。

5 应用

为说明我们的框架，我们将结果应用于若干类广受关注的多指标模型。

5.1 学习高斯多指标模型

考虑一个链接函数为 $\rho \in \mathcal{P}(\mathcal{Y}\times\mathbb{R}^s)$、支撑集为 $\boldsymbol{W}_* \in \mathsf{Stf}_s(\mathbb{R}^d)$ 的高斯多指标模型：

$$(y,\boldsymbol{x}) \sim \mathbb{P}_\rho^{\boldsymbol{W}_*}: \qquad \boldsymbol{x}\sim\gamma_d, \qquad y \mid \boldsymbol{x} \sim \rho(\cdot \mid \boldsymbol{W}_*^{\mathsf{T}}\boldsymbol{x}). \tag{49}$$

这一设定已得到广泛研究；例如参见 [DH18、BKM+19、MM18、TDD+24、DPVLB24、DLB25、DIKR25、DIKZ25] 及其中的参考文献。特别地，弱恢复（在局部差分隐私框架内）的最优样本复杂度为 $n \asymp d^{k_*/2}$，其中 k_* 是生成指数，其定义为 ρ [DPVLB24] 的埃尔米特展开中第一个非零系数的指标。与我们的框架类似，完整子空间 $\boldsymbol{W}_*$ 的恢复随后通过一个迭代过程实现：该过程反复对当前已恢复子空间进行条件化 [DLB25、DIKR25、DIKZ25]。

我们使用本框架重新审视这一经典设定。如例1所述，该模型可通过极分解 $\boldsymbol{x} = r\boldsymbol{z}$重写为球面 MIM，其中 $r = \|\boldsymbol{x}\|_2 \sim \chi_d$ 和 $\boldsymbol{z} = \boldsymbol{x}/\|\boldsymbol{x}\|_2 \sim \tau_d$ 相互独立。为清晰起见，我们重点分析单次学习步骤的复杂度；相同的分析可自然扩展至完整的迭代过程。准确的表述、证明及更多细节见附录 C.1。

Let

$$\boldsymbol{\psi}_{\emptyset,k}(y) := \mathbb{E}[\mathrm{He}_k(\boldsymbol{x}) \mid y] \in \mathrm{Sym}_k(\mathbb{R}^d)$$

表示阶数为k 的埃尔米特系数，其中 ρ，$\mathrm{He}_k(\boldsymbol{x})$ $\in\mathrm{Sym}_k(\mathbb{R}^d)$ 是埃尔米特张量（见附录 74 中的公式 C.1）。生成指数为

$$k_* = \arg\min_{k\ge 1}\{k \ : \ \|\boldsymbol{\psi}_{\emptyset,k}\|_{L^2(\rho^Y)}^2 > 0\},$$

其中 ρ^Y 表示 y 的边缘分布。相比之下，定理1 和3 表明，计算复杂度由调和分解决定，尤其取决于$\|\boldsymbol{\xi}_{\emptyset,\ \ell}\|_{L^2}$当 $\ell \geq 1$ 时，其中

$$\boldsymbol{\xi}_{\emptyset,\ell}(y,r) := \mathbb{E}[\mathcal{H}_{d,\ell}(\boldsymbol{z}) \mid y,r] \in \mathrm{TSym}_\ell(\mathbb{R}^d)$$

是调和系数。为建立调和系数与埃尔米特系数之间的联系，我们将埃尔米特张量分解为调和张量：对所有 $\boldsymbol{A} \in \mathrm{Sym}_k(\mathbb{R}^d)$,

$$\langle \boldsymbol{A}, \mathrm{He}_k(\boldsymbol{x})\rangle_{\mathsf{F}} = \sum_{j=0}^{\lfloor k/2\rfloor} \beta_{k,k-2j}^{(d)}(r)\, \langle \mathsf{P}_{\mathsf{tf}}\left(\tau^j(\boldsymbol{A})\right), \mathcal{H}_{d,k-2j}(\boldsymbol{z})\rangle_{\mathsf{F}}, \tag{50}$$

其中 $\|\beta_{\ell+2j,\ \ell}^{(d)}\|_{L^2}^2 \asymp d^{-j}$ （显式表达式见附录 C.1 中的引理 9 ）。在对 ρ 作出适当正则性假设的条件下，该分解使我们能够将$\|\boldsymbol{\xi}_{\emptyset,\ \ell}\|_{L^2}$ 与$\|\boldsymbol{\psi}_{\emptyset,\ \ell}\|_{L^2}$联系起来。

首先，$\|\boldsymbol{\xi}_{\emptyset,\ k_*}\|_{L^2}^2 \asymp \|\boldsymbol{\psi}_{\emptyset,\ k_*}\|_{L^2}^2 \asymp 1$。因此，在次数 $\ell = k_*$ 处应用张量展开可得

$$n \asymp d^{k_*/2 \vee 1}, \qquad \mathtt{T} \asymp d^{k_* \vee (3/2) + \delta_{k_*}/2} \log(d), \tag{51}$$

达到与 [DLB25]相匹配的最优样本复杂度。运行时间上界由形状为 $(a,\ b) = (\lfloor k_*/2\rfloor,\ \lceil k_*/2\rceil)$ 的张量展开得出，并且只要满足 $k_* > 2$，就优于 [DLB25]。事实上， [DLB25] 依赖于矩形张量展开，其效果等同于在我们的框架中选择 $(a,\ b) = (1,\ k_*-1)$，导致运行时间为 $\mathtt{T} \asymp d^{3k_*/2-1}\log d$。因此，尽管矩形展开足以实现最优样本复杂度，但更加均衡的重塑在计算上更高效。

虽然（51）实现了最优样本复杂度，但次数 k_* 不一定是运行时间最优的。事实上，根据（50），埃尔米特次数 k 会贡献于调和次数 $\ell < k$。因此，对于$\ell < k_*$ ，只要埃尔米特系数的偏迹非零，$\boldsymbol{\xi}_{\emptyset,\ \ell}$ 就可能非零。例如，令 j_* 为满足 $\|\tau^j(\boldsymbol{\psi}_{\emptyset,\ k_*})\|_{L^2} > 0$的最大整数 $0 \leq j < \lceil k_*/2\rceil$ 。那么$\|\boldsymbol{\xi}_{\emptyset,\ k_*-2j_*}\|_{L^2} \asymp d^{-j_*}$，并在次数 $\ell_* = k_* - 2j_*$ 处应用调和张量展开（为简化起见，假设 $\ell_* \geq 2$ ）可得

$$n \asymp d^{k_*/2}, \qquad \mathtt{T} \asymp d^{k_* - j_* + \frac{1}{2}\delta_{\ell_*}} \log(d).$$

考虑第1.2.3节介绍的以下两个例子：

- 对于s-奇偶函数（13），可以证明存在某个 $h\colon\ \mathcal{Y} \to \mathbb{R}$,

  $$\boldsymbol{\psi}_{\emptyset,\mathsf{s}}(y) = h(y)\, \mathsf{P}_{\mathsf{sym}}(\boldsymbol{w}_{*,1} \otimes \cdots \otimes \boldsymbol{w}_{*,\mathsf{s}}),$$

  且所有部分 迹均消失。在这种情况下， $\ell = \mathsf{s}$ 既是样本最优的，也是运行时间最优的 .

- 对于生成指数为 k_*的高斯单指标模型（12），

  $$\boldsymbol{\psi}_{\emptyset,k_*}(y) = h(y)\, \boldsymbol{w}_*^{\otimes k_*}, \qquad \tau^j(\boldsymbol{\psi}_{\emptyset,k_*}(y)) = h(y)\, \boldsymbol{w}_*^{\otimes(k_*-2j)}.$$

  尽可能多地进行收缩，可将次数降至 $\ell_* = 1$ 或 2。样本复杂度仍为 $n \asymp d^{k_*/2}$，而（针对奇数 k_* [JKMS25]修改算法后）

  $$\mathtt{T} \asymp d^{k_\star/2+1}\log(d).$$

  这恢复了 [DPVLB24]的偏迹估计量。

在上述例子中，可以在不牺牲最优样本复杂度的情况下，实现接近最优的运行时间。但更一般地，系数$\psi_{\emptyset,\, k}$在 $k > k_*$ 时，可能会产生样本复杂度与运行时间的权衡。在这种情况下，若要提高运行时间，就必须接受更差的样本复杂度；这与高斯单指标模型不同，后者总能同时达到最优。为说明这一点，固定一个整数 $p \geq 1$，并考虑由一个 $4p$-奇偶函数和一个生成指数为 $6p$ 的高斯单指标模型构成的混合分布，二者出现的概率相等。由于 $4p$-奇偶函数这一分量，样本最优次数为 $\ell = 4p$，因此得到

$$n \asymp d^{2p}, \qquad \mathtt{T} \asymp d^{4p} \log(d).$$

运行　　时间最优次数为 $\ell = 2$，这得益于高斯单指标模型分量，从而得到

$$n \asymp d^{3p}, \qquad \mathtt{T} \asymp d^{3p+1} \log(d).$$

附录 C.1.2 根据 ρ 的埃尔米特系数，对样本复杂度和运行时间最优次数及其相关复杂度给出了通用刻画。

总之，对于高斯多指标模型，最优样本复杂度仅取决于生成指数，而计算复杂度则可能取决于更高次的埃尔米特系数及其偏迹。我们的调和分析框架自然且简洁地刻画了这一点。

5.2 学习方向多指标模型

考虑与上一节相同的高斯多指标模型（49），但现在假设径向分量 r 不可观测，且只能获得（y, $\boldsymbol{z}$）（即标签和输入的方向）。我们将这种模型称为方向多指标模型。这种设定源于统计学和机器学习中的一种常见做法：将输入向量归一化为固定范数。尽管由此得到的模型不再是高斯多指标模型，但它仍是一个球面多指标模型（定义在$\mathcal{Y} \times \mathbb{S}^{d-1}$上），因此可以在我们的框架内进行分析。

与上一节一样，我们使用厄米特-调和分解。然而，在方向情形下，相关量按 $\mathbb{E}[\beta^{(d)}_{\ell+2j,\ell}(r)]^2 \asymp d^{-2j}$ 的尺度变化（而 $\mathbb{E}[\beta^{(d)}_{\ell+2j,\ell}(r)^2] \asymp d^{-j}$）。因此，选择调和次数 $\ell < k_*$ 总会增加样本复杂度，却不会改善运行时间。因此，与高斯多指标模型情形不同，方向 MIM 中不存在样本复杂度与运行时间复杂度之间的权衡。最优选择始终是 $\ell = k_*$，而该次数下的张量展开给出

$$n \asymp d^{k_*/2}, \qquad \mathtt{T} \asymp d^{k_* + \frac{1}{2}\delta_{k_*}} \log(d).$$

特别是，尽管最优样本复杂度保持不变，但仅有方向信息可用时（即输入已归一化），运行时间可能会显著变差。例如，在高斯单指标模型中，运行时间从 $d^{k_*/2+1}$ 增加到 d^{k_*}。此前，[JKMS25], 已在高斯单指标模型中观察到这一现象，并证明它对基于梯度的算法和算法设计有重要影响。

完整的陈述和证明请参见附录 C.1.3 。

5.3 其他应用

另一个自然的研究方向是考察具有结构化链接函数的球面 MIM。考虑输入 $\boldsymbol{x} \sim \mu$，其具有球面对称不变性，且极分解为 $\boldsymbol{x} = r\boldsymbol{z}$，其中$r \sim \mu_r$ 且 $\boldsymbol{z} \sim \tau_d$。令 $\boldsymbol{W}_* \in \mathsf{St}_{\mathsf{s}}(\mathbb{R}^d)$ 表示支撑集，并考虑如下形式的响应：

$$y = f(\boldsymbol{W}_*^{\mathsf{T}} \boldsymbol{x}) + \epsilon,$$

其中 ϵ 是独立噪声，且 $f:\ \mathbb{R}^s \to \mathbb{R}$ 属于一个预先指定的函数类。相关例子包括多项式、半空间交集、分段线性函数（例如 ReLU 网络或多类线性分类）、多层神经网络以及奇偶函数。在高斯情形下，此类函数类的学习已在 [DLB25、DIKR25、DIKZ25]及其中的参考文献中得到研究。

我们的框架可以为一般的球面对称不变输入分布给出类似的保证。作为一个简单例子，可以直接分析奇偶函数$f(\boldsymbol{t}) = \mathsf{sign}(t_1 t_2 \cdots t_s)$的调和分解（它不依赖于径向分量 r），如第1.2.3节所述。半空间交集 $f(\boldsymbol{t}) = 2\prod_{j=1}^{s} \mathbb{1}(\langle \boldsymbol{a}_j, \boldsymbol{t}\rangle \geq b_j) - 1$可以（在对 μ_r的适当正则性假设下）通过改写 [DLB25,DIKZ25]的论证来研究。作为另一个例子，考虑多项式链接函数 $f(\boldsymbol{t}) = \sum_{|\alpha|\leq D} c_\alpha \boldsymbol{t}^\alpha$，其中次数 $D \in \mathbb{N}$固定。在附录 C.2中，我们证明当$\mu_r = \delta_{\sqrt{d}}$时，以下结论成立：对于任意维数为 $s_{\boldsymbol{U}} < s$的已恢复子空间 $\boldsymbol{U}$，有$\|\boldsymbol{\xi}_{\boldsymbol{U},2}\|_{L^2}^2 \asymp 1$. 因此，在每一步以次数 $\ell = 2$ 应用迭代张量展开，即可恢复植入子空间，且

$$n \asymp d, \qquad \mathtt{T} \asymp d^2 \log(d).$$

不出所料，这与高斯情形下的复杂度一致，其中 [DLB25]表明生成跃迁复杂度至多为 2。

我们将对更多函数类的探索留待未来工作。

6 讨论

在本文中，我们引入了球面多指标模型，即流行的高斯多指标模型的一种自然等变扩展，并刻画了在局部差分隐私与统计查询框架下支撑恢复的样本复杂度与运行时间复杂度。我们进一步提出了一族基于调和张量展开的迭代算法，其（几乎）匹配这些下界。具体而言，这些算法可通过选择恢复过程中所使用的调和次数序列，在样本复杂度与运行时间复杂度之间实现不同的权衡。我们的刻画依赖于将学习问题分解为 $\mathcal{O}_d$ 的不可约表示，并利用其与无迹对称张量之间的交织同构。作为应用，我们重新考察高斯多指标模型：除了恢复 [DLB25，DIKR25],中所建立的最优样本复杂度之外，我们还得到了改进的运行时间上界。具体而言，我们给出的算法其查询复杂度紧密贴近统计查询下界，尽管在某些区域这可能以牺牲最优样本复杂度为代价。

局限性。 我们的工作存在一些局限性。首先，为了简化，我们假设 ν_d 是固定且已知的。尽管我们预计这一假设可以在我们的算法中放宽，但我们将这一扩展留待未来工作。其次，我们的下界针对检测而表述。特别是，对于 $\ell = 1$，可能存在检测-恢复差距。要解决这一问题，需要直接针对估计证明低次下界；近期在这一方向上的进展包括 [SW22、CGGV25]。第三，我们的下界依赖于低次猜想以及与查询复杂度之间的启发式对应关系；例如，参见 [Wei25、DKLP25、CMZZ25]。

未来研究方向。 未来研究有若干自然的方向，包括：(i) 分析用于学习多指标模型的基于梯度的方法（例如，参见 [BBPV25，MW26]）；(ii) 利用高阶方法刻画固定调和子空间内更精细的样本量与运行时间的权衡

张量表示（见注 3.3）；(iii) 理解球面对称性被破坏时的学习问题；以及 (iv) 研究秩 s随 d增长的区域（例如，参见 [OSSW24]）。

最后，我们的上下界依赖于正交群的一般性质 $\mathcal{O}_d$及其不可约表示。因此，同样的分析可直接应用于其他 $\mathcal{O}_d$-等变模型，包括单尖峰和多尖峰张量主成分分析。特别是，张量展开[MR14] 和偏迹估计量 [HSSS16]可解释为作用于特定不可约分量 $\mathrm{TSym}^\ell(\mathbb{R}^d)$，其中 $\ell = k$（张量阶数）和 $\ell \in \{1, 2\}$ 分别表示相应阶数。更广泛地说，在后续论文 [JKMS26], 中，我们将其中一些思想拓展到一般的等变学习问题和紧群作用。这表明，有望基于表示论原理，系统地构建统计最优且计算最优的等变学习算法理论。

参考文献

[AAM22] Emmanuel Abbe、Enric Boix Adsera 和 Theodor Misiakiewicz，《合并阶梯性质：双层神经网络中随机梯度下降学习稀疏函数的必要且近乎充分条件》，学习理论会议，PMLR，2022 年，第 4782–4887 页。[AAM23] ,《神经网络上的随机梯度下降学习：跃迁复杂度与鞍点到鞍点动力学》，第三十六届学习理论年会，PMLR，2023 年，第 2552–2623 页。[ABAB+21] Emmanuel Abbe、Enric Boix-Adsera、Matthew S Brennan、Guy Bresler 和 Dheeraj Nagaraj，《阶梯性质：层次结构如何引导深度学习》，神经信息处理系统进展《34》（2021），26989–27002。[Abb18] Emmanuel Abbe，《社区检测与随机块模型：近期进展》，《机器学习研究》《18》（2018），第 177 期，1–86。[ADK+24a] Luca Arnaboldi、Yatin Dandi、Florent Krzakala、Bruno Loureiro、Luca Pesce 和 Ludovic Stephan，《在线学习与信息指数：批大小及时间与复杂度权衡的重要性》，国际机器学习会议，PMLR，2024 年，第 1730–1762 页。[ADK+24b] Luca Arnaboldi、Yatin Dandi、Florent Krzakala、Luca Pesce 和 Ludovic Stephan，《温故知新：数据重复使随机梯度下降能够学习高维多指标函数》，arXiv 预印本 arXiv:2405.15459（2024）。[AGJ21] Gerard Ben Arous、Reza Gheissari 和 Aukosh Jagannath，《高维推断中非凸损失上的在线随机梯度下降》，《机器学习研究》《22》（2021），第 106 期，1–51。[AKS98] Noga Alon、Michael Krivelevich 和 Benny Sudakov，《在随机图中寻找大型隐藏团》，《随机结构与算法》《13》（1998），第 3–4 期，457–466。[ALPV14] Albert Ai、Alex Lapanowski、Yaniv Plan 和 Roman Vershynin，《非高斯测量下的一比特压缩感知》，《线性代数及其应用》《441》（2014），222–239。

[ASKL23] Luca Arnaboldi、Ludovic Stephan、Florent Krzakala 和 Bruno Loureiro，《从高维与平均场动力学到无量纲常微分方程：双层网络随机梯度下降的统一方法》，第三十六届学习理论年会，PMLR，2023 年，第 1199–1227 页。[Bau90] Eric B Baum，《关于学习半空间的并集》，《复杂性杂志》6（1990），第 1 期，67–101。[BB18] Dmitry Babichev 和 Francis Bach，《利用得分函数进行切片逆回归》，《电子统计学杂志》12（2018），第 1 期，1507–1543。[BB19] Matthew Brennan 和 Guy Bresler，《稀疏主成分分析的最优平均情形归约：从弱假设到强困难性》，学习理论会议，PMLR，2019 年，第 469–470 页。[BBPV25] Alberto Bietti、Joan Bruna 和 Loucas Pillaud-Vivien，《梯度流学习高斯多指标模型（第一部分）：一般性质与双时间尺度学习》，《纯粹与应用数学通讯》78（2025），第 12 期，2354–2435。[BBvH23] Afonso S. Bandeira、March T. Boedihardjo 和 Ramon van Handel，《矩阵集中不等式与自由概率》，《数学新进展》234（2023），第 1 期，419–487（英文）。[BEAH+22] Afonso S Bandeira、Ahmed El Alaoui、Samuel Hopkins、Tselil Schramm、Alexander S Wein 和 Ilias Zadik，《Franz-Parisi 判据与高维统计中的计算权衡》，神经信息处理系统进展 35（2022），33831–33844。[Bec92] William Beckner，《Sobolev不等式、泊松半群与球面分析》 s^n。，美国国家科学院院刊 89（1992），第 11 期，4816–4819。[BGL13] Dominique Bakry、Ivan Gentil 和 Michel Ledoux，《马尔可夫扩散算子的分析与几何》，第 348 卷，施普林格科学与商业媒体，2013 年。[BH25] Joan Bruna 和 Daniel Hsu，《多指标模型算法综述》，《统计科学》40（2025），第 3 期，378–391。[Bha97] Rajendra Bhatia，《正规矩阵谱子空间的扰动》，第 194–225 页，施普林格，纽约，纽约州，1997 年（英文）。[BHK+19] Boaz Barak、Samuel Hopkins、Jonathan Kelner、Pravesh K Kothari、Ankur Moitra 和 Aaron Potechin，《植入团问题的近乎紧致平方和下界》，《SIAM计算杂志》48（2019），第 2 期，687–735。[BJW19] Ainesh Bakshi、Rajesh Jayaram 和 David P Woodruff，《在多项式时间内学习双层整流神经网络》，学习理论会议，PMLR，2019 年，第 195–268 页。[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（2019），第 12 期，5451–5460。

[BLM13] St´ephane Boucheron、G´abor Lugosi 和 Pascal Massart，《集中不等式：独立性的非渐近理论》，牛津大学出版社，第 2 版，2013 年。[BR13] Quentin Berthet 和 Philippe Rigollet，《稀疏主成分检测的复杂性理论下界》，学习理论会议，PMLR，2013 年，第 1046–1066 页。[Bri82] David R Brillinger，《具有“高斯”回归变量的广义线性模型》，《Erich L. Lehmann 纪念文集》（1982），第 97 页。[Buc69] Herbert Buchholz，《涉及抛物函数的广义积分和定积分及若干无穷级数》，第 112–134 页，施普林格，柏林、海德堡，1969 年（英文）。[BvH24] Tatiana Brailovskaya 和 Ramon van Handel，《普适性与尖锐的矩阵集中不等式》，《几何与泛函分析》34（2024），第 6 期，1734–1838。[CDG+23] Sitan Chen、Zehao Dou、Surbhi Goel、Adam Klivans 和 Raghu Meka，《学习窄型单隐藏层 ReLU 网络》，第三十六届学习理论年会，PMLR，2023 年，第 5580–5614 页。[CGGV25] Alexandra Carpentier、Simone Maria Giancola、Christophe Giraud 和 Nicolas Verzelen，《通过近似正交归一基得到的低次下界》，arXiv:2509.09353（2025）。[Chi11] Theodore S Chihara，《正交多项式导论》，Courier Corporation，2011 年。[CKM22] Sitan Chen、Adam R Klivans 和 Raghu Meka，《学习深层 ReLU 网络是固定参数可处理的》，2021 年 IEEE 第 62 届计算机科学基础年会（FOCS），IEEE，2022 年，第 696–707 页。[CLS15] Emmanuel J Candes、Xiaodong Li 和 Mahdi Soltanolkotabi，《通过 Wirtinger 流进行相位恢复：理论与算法》，《 IEEE 信息论汇刊》61（2015），第 4 期，1985–2007。[CM20] Sitan Chen 和 Raghu Meka，《在少数相关维度中学习多项式》，学习理论会议，PMLR，2020 年，第 1161–1227 页。[CMZZ25] Siyu Chen、Theodor Misiakiewicz、Ilias Zadik 和 Peiyuan Zhang，《优化的 Franz-Parisi 判据及其与统计查询下界的等价性》，arXiv:2506.06259（2025）。[CSV13] Emmanuel J Candes、Thomas Strohmer 和 Vladislav Voroninski，《PhaseLift：通过凸规划从幅度测量中精确、稳定地恢复信号》，《纯粹与应用数学通讯》66（2013），第 8 期，1241–1274。[DC00] R Dennis Cook，《SAVE：一种用于回归降维和图形分析的方法》，《统计学通讯：理论与方法》29（2000），第 9–10 期，2109–2121。[DDM+25] Leonardo Defilippis、Yatin Dandi、Pierre Mergny、Florent Krzakala 和 Bruno Loureiro，《高维多指标模型中的最优谱转变》，arXiv 预印本 arXiv:2502.02545（2025）。[DH18] Rishabh Dudeja 和 Daniel Hsu，《在高斯空间中学习单指标模型》，学习理论会议，PMLR，2018 年，第 1887–1930 页。

[DIKR25] Ilias Diakonikolas、Giannis Iakovidis、Daniel Kane 和 Lisheng Ren，《鲁棒学习实值多指标模型的算法与统计查询下界》，第三十九届神经信息处理系统年会，2025 年。[DIKZ25] Ilias Diakonikolas、Giannis Iakovidis、Daniel M Kane 和 Nikos Zarifis，《通过迭代子空间近似鲁棒学习多指标模型》，arXiv:2502.09525（2025）。[DJS08] Arnak S Dalalyan、Anatoly Juditsky 和 Vladimir Spokoiny，《一种估计有效降维子空间的新算法》，《机器学习研究》9（2008），1647–1678。[DKLP25] Ilias Diakonikolas、Daniel M Kane、Sihan Liu 和 Thanasis Pittas，《非高斯成分分析的 Ptf 检验下界》，arXiv:2511.19398（2025）。[DKMR22] Ilias Diakonikolas、Daniel Kane、Pasin Manurangsi 和 Lisheng Ren，《对抗性标签噪声下学习单神经元的计算困难性》，国际人工智能与统计会议，PMLR，2022 年，第 8199–8213 页。[DKR23] Ilias Diakonikolas、Daniel Kane 和 Lisheng Ren，《在高斯边缘分布下不可知地学习半空间和 ReLU 回归的近最优密码学困难性》，国际机器学习会议，PMLR，2023 年，第 7922–7938 页。[DKTZ22] Ilias Diakonikolas、Vasilis Kontonis、Christos Tzamos 和 Nikos Zarifis，《通过梯度下降在对抗性标签噪声下学习单神经元》，学习理论会议，PMLR，2022 年，第 4313–4361 页。[DLB25] Alex Damian、Jason D Lee 和 Joan Bruna，《生成式飞跃：高效学习高斯多指标模型的紧样本复杂度》，第三十九届神经信息处理系统年会，2025 年。[dlPMS99] Victor H. de la Pe˜na 和 Stephen J. Montgomery-Smith，《多元 U 统计量尾概率的解耦不等式》，1999 年。[DLS22] Alexandru Damian、Jason Lee 和 Mahdi Soltanolkotabi，《神经网络可以通过梯度下降学习表示》，学习理论会议，PMLR，2022 年，第 5413–5452 页。[DNGL24] Alex Damian、Eshaan Nichani、Rong Ge 和 Jason D Lee，《平滑优化景观可增强随机梯度下降的信号：学习单指标模型的最优样本复杂度》，神经信息处理系统进展第36卷（2024），第 1–39 页。[DPVLB24] Alex Damian、Loucas Pillaud-Vivien、Jason Lee 和 Joan Bruna，《高斯单指标模型中的计算-统计差距》，第三十七届学习理论年会，PMLR，2024 年，第 1262–1262 页。[DTA+24] Yatin Dandi、Emanuele Troiani、Luca Arnaboldi、Luca Pesce、Lenka Zdeborov´a 和 Florent Krzakala，《在两层网络的梯度下降中重复使用批次的益处：打破信息诅咒与飞跃指数》，arXiv 预印本 arXiv:2402.03220（2024），第 1–30 页。[DV21] Amit Daniely 和 Gal Vardi，《从局部伪随机生成器到学习的计算困难性》，学习理论会议，PMLR，2021 年，第 1358–1394 页。

[Far08] Jacques Faraut，《李群分析：导论》，剑桥高等数学研究丛书，剑桥大学出版社，2008。[Fel07] Vitaly Feldman，奇偶函数和 DNF 表达式的属性高效与非自适应学习，《机器学习研究》8（2007），第7期。[Fel17] ，统计查询复杂度的一般刻画，学习理论会议，PMLR，2017，第785–830页。[FGR+17] Vitaly Feldman、Elena Grigorescu、Lev Reyzin、Santosh S Vempala 和 Ying Xiao，统计算法与检测植入团的下界，ACM期刊（JACM）64（2017），第2期，1–37。[FH13] William Fulton 和 Joe Harris，《表示论：入门课程》，第129卷，施普林格科学与商业媒体，2013。.[Fis18] Ernst Fischer，《关于代数的微分过程》，Journal f¨ur die reine und angewandte Mathematik（Crelles Journal）1918（1918），第148期，1–78。[GGKS23] Aravind Gollakota、Parikshit Gopalan、Adam Klivans 和 Konstantinos Stavropoulos，使用全能预测器进行单指标模型的无偏学习，神经信息处理系统进展第36卷（2023），14685–14704。[GLM18] Rong Ge、Jason D Lee 和 Tengyu Ma，通过景观设计学习单隐层神经网络，国际学习表征会议，2018。[GLPTJ17] Olivier Gu´edon、Alexander E. Litvak、Alain Pajor 和 Nicole Tomczak-Jaegermann，随机矩阵奇异值的波动区间，欧洲数学会杂志19（2017），第5期，1469–1505（英文）。[GNJN13] Sivakant Gopi、Praneeth Netrapalli、Prateek Jain 和 Aditya Nori，一位压缩感知：可证明的支撑集与向量恢复，国际机器学习会议，PMLR，2013，第154–162页。[GW00] Roe Goodman 和 Nolan R Wallach，经典群的表示与不变量，剑桥大学出版社，2000。[GW09] Roe Goodman 和 Nolan R. Wallach，《对称性、表示与不变量》，数学研究生教材，施普林格，纽约，纽约州，2009（英文）。[HJS01] Marian Hristache、Anatoli Juditsky 和 Vladimir Spokoiny，单指标模型中指标系数的直接估计，统计年刊（2001），595–623。[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，2017 IEEE第58届计算机科学基础年会（FOCS），IEEE，2017，第720–731页。[Hop18] Samuel Hopkins，统计推断与平方和方法，博士论文，康奈尔大学，2018。[Hor09] Joel L. Horowitz，单指标模型，第7–51页，美国施普林格，纽约，纽约州，2009（英文）。

[HS89] Wolfgang H¨ardle 和 Thomas M Stoker，使用平均导数法研究光滑多元回归，美国统计学会期刊84（1989），第408期，986–995。[HSSS16] Samuel B Hopkins、Tselil Schramm、Jonathan Shi 和 David Steurer，基于平方和证明的快速谱算法：张量分解与植入稀疏向量，第四十八届ACM计算理论研讨会论文集，2016，第178–191页。[HWX15] Bruce Hajek、Yihong Wu 和 Jiaming Xu，随机图社区检测的计算下界，学习理论会议，PMLR，2015，第899–928页。[JKMS25] Nirmit Joshi、Hugo Koubbi、Theodor Misiakiewicz 和 Nathan Srebro，通过调和分解学习单指标模型，第三十九届神经信息处理系统年会，2025。[JKMS26] Nirmit Joshi、Hugo Koubbi、Theodor Misiakiewicz 和 Nathan Srebro，群等变性下的学习：调和分解、群分解与跃迁复杂度，准备中（2026）。[JMS24] Nirmit Joshi、Theodor Misiakiewicz 和 Nati Srebro，关于使用统计查询和梯度查询学习稀疏函数的复杂度，神经信息处理系统进展第37卷（2024），103198–103241。[Kea98] Michael Kearns，基于统计查询的高效抗噪学习，ACM期刊（JACM）45（1998），第6期，983–1006。[KKSK11] Sham M Kakade、Varun Kanade、Ohad Shamir 和 Adam Kalai，利用保序回归高效学习广义线性模型和单指标模型，神经信息处理系统进展第24卷（2011），1–17。[KLV21] Timo Klock、Alessandro Lanteri 和 Stefano Vigogna，使用响应条件最小二乘法估计多指标模型，《电子统计学杂志》15（2021），589–629。[KS09a] Adam Tauman Kalai 和 Ravi Sastry，isotron算法：高维保序回归，COLT，第1卷，2009，第9页。[KS09b] Adam R Klivans 和 Alexander A Sherstov，学习半空间交集的密码学困难性，计算机与系统科学杂志75（2009），第1期，2–12。[KSV24] Adam Klivans、Konstantinos Stavropoulos 和 Arsen Vasilyan，分布偏移下半空间交集的学习：改进的算法与统计查询下界，第三十七届学习理论年会，PMLR，2024，第2944–2978页。[KWB19] D mitriy Kunisky、Alexander S Wein 和 Afonso S Bandeira，关于假设检验计算困难性的笔记：基于低次似然比的预测，ISAAC大会（国际分析、应用分析与计算学会），施普林格，2019，第1–50页。

[KX26] Pravesh K Kothari 和 Jeff Xu，基于 Kikuchi 矩阵的锐利界限研究张量主成分分析的平滑权衡，《2026 年 ACM-SIAM 离散算法研讨会》（SODA）论文集，SIAM，2026，第 2617–2632 页。[KYM25] Filip Kovaˇcevi´c、Zhang Yihan 和 Marco Mondelli，多指标模型的谱估计器：精确渐近行为与最优弱恢复，《第三十八届学习理论会议》论文集（Nika Haghtalab 和 Ankur Moitra 编），机器学习研究论文集，第 291 卷，PMLR，2025，第 3354–3404 页。[Li91] Ker-Chau Li，切片逆回归与降维，《美国统计学会期刊》86（1991），第 414 期，316–327 页。[Li92] ,关于用于数据可视化和降维的主海森方向：Stein 引理的另一种应用，《美国统计学会期刊》87（1992），第 420 期，1025–1039 页。[Li25] Zhangsong Li，张量主成分分析中的平滑计算转变，arXiv:2509.09904（2025）。[LKDD24] Shuyao Li、Sushrut Karmalkar、Ilias Diakonikolas 和 Jelena Diakonikolas，学习单神经元，使其能够稳健应对分布偏移和对抗性标签噪声，《神经信息处理系统进展》第37卷（2024），67383–67421 页。[LL20] Yue M Lu 和 Gen Li，高维非凸估计中谱初始化的相变，《信息与推断：IMA 期刊》9（2020），第 3 期，507–541 页。[LOSW24] Jason D Lee、Kazusato Oko、Taiji Suzuki 和 Denny Wu，神经网络在信息论极限附近通过随机梯度下降学习低维多项式，《神经信息处理系统进展》第37卷（2024），58716–58756 页。[LW07] Bing Li 和 Shaoli Wang，关于用于降维的方向回归，《美国统计学会期刊》102（2007），第 479 期，997–1008 页。[Man94] Yishay Mansour，通过傅里叶变换学习布尔函数，《神经计算与学习理论进展》，施普林格，1994，第 391–424 页。[McC84] Peter McCullagh，广义线性模型，《欧洲运筹学杂志》16（1984），第 3 期，285–292 页。[Mis25] Theodor Misiakiewicz，深度学习计算复杂度简明教程。[MLKZ20] Antoine Maillard、Bruno Loureiro、Florent Krzakala 和 Lenka Zdeborov´a，高维相位恢复：统计与计算相变，《神经信息处理系统进展》33（2020），11071–11082 页。[MM18] Marco Mondelli 和 Andrea Montanari，弱恢复的基本极限及其在相位恢复中的应用，《学习理论会议》，PMLR，2018，第 1445–1450 页。[MR14] Andrea Montanari 和 Emile Richard，张量主成分分析的统计模型，《神经信息处理系统进展》27（2014），1–30 页。

[MW26] Andrea Montanari 和 Zihao Wang，神经网络特征学习的相变，arXiv:2602.01434（2026）。[MZST06] Sayan Mukherjee、Ding-Xuan Zhou 和 John Shawe-Taylor，通过梯度学习坐标协方差。《机器学习研究》7（2006），第 3 期。[NJS13] Praneeth Netrapalli、Prateek Jain 和 Sujay Sanghavi，通过交替最小化进行相位恢复，《神经信息处理系统进展》26（2013）。[NW72] J. A. Nelder 和 R. W. M. Wedderburn，广义线性模型，《皇家统计学会会刊》A辑（总论）135（1972），第 3 期，370–384 页。[OSSW24] Kazusato Oko、Yujin Song、Taiji Suzuki 和 Denny Wu，学习多样特征之和：岭组合的计算困难性与高效的基于梯度的训练，《第三十七届学习理论年会》，PMLR，2024，第 4009–4081 页。[PV13] Yaniv Plan 和 Roman Vershynin，通过线性规划进行一比特压缩感知，《纯粹与应用数学通讯》 66 （ 2013 ），第 8 期，1275–1297 页（英文）。[Rey20] Lev Reyzin，统计查询与统计算法：基础与应用，arXiv 预印本 arXiv:2004.00557（2020），1–21 页。[Rot86] O.S Rothaus，紧李群上的超收缩性与 Bakry-Émery 准则，《泛函分析杂志》65（1986），第 3 期，358–367 页。[Sam93] Alexander M Samarov，通过非参数函数估计探索回归结构，《美国统计学会期刊》88（1993），第 423 期，836–847 页。[SSBD14] Shai Shalev-Shwartz 和 Shai Ben-David，《理解机器学习：从理论到算法》，剑桥大学出版社，2014。[SW22] Tselil Schramm 和 Alexander S Wein，低次多项式估计的计算障碍，《统计学年刊》50（2022），第 3 期，1833–1858 页。[SZB21] Min Jae Song、Ilias Zadik 和 Joan Bruna，学习单周期神经元的密码学困难性，《神经信息处理系统进展》34（2021），29602–29615 页。[Sze39] Gabor Szeg，正交多项式，第 23 卷，美国数学学会，1939。[TDD+24] Emanuele Troiani、Yatin Dandi、Leonardo Defilippis、Lenka Zdeborov´a、Bruno Loureiro 和 Florent Krzakala，高维多指标模型中弱可学习性的基本极限，《高维学习动力学 2024：结构与推理的涌现》，2024。[TWKS14] Shubhendu Trivedi、Jialei Wang、Samory Kpotufe 和 Gregory Shakhnarovich，期望梯度外积的一致估计量。UAI，2014，第 819–828 页。[Vem97] Santosh Vempala，一种基于随机抽样、用于学习半空间交集的算法，《第38届计算机科学基础年会论文集》，IEEE，1997，第 508–513 页。

[Vem10] Santosh S Vempala，利用主成分分析从高斯分布中学习凸概念，2010年IEEE第51届计算机科学基础年会，IEEE，2010，第124–130页。[WEAM19] Alexander S Wein、Ahmed El Alaoui 和 Cristopher Moore，*Kikuchi*层级与张量主成分分析，2019年IEEE第60届计算机科学基础年会（FOCS），IEEE，2019，第1446–1468页。[Wei25] Alexander S Wein，统计学的计算复杂性：低次多项式带来的新见解，arXiv:2506.10748（2025）。[WZDD24] Puqian Wang、Nikos Zarifis、Ilias Diakonikolas 和 Jelena Diakonikolas，样本与计算高效地鲁棒学习高斯单指标模型，神经信息处理系统进展**37**（2024），58376–58422。[ZPVB23] Aaron Zweig、Loucas Pillaud-Vivien 和 Joan Bruna，超越高斯数据的单指标模型，神经信息处理系统进展**36**（2023），10210–10222。[ZWDD24] Nikos Zarifis、Puqian Wang、Ilias Diakonikolas 和 Jelena Diakonikolas，通过对齐锐度鲁棒学习单指标模型，arXiv预印本 arXiv:2402.17756（2024）。

下界的证明

注 A.1。 查询对齐和查询跃迁复杂度中出现的算子范数 $\|\mathbf{\Gamma}_{\boldsymbol{U},\ \ell}\|_{\mathrm{op}}$ 具有自然的变分解释：它们刻画了广义响应 $\boldsymbol{y}_{\boldsymbol{U}}$ 的非线性变换与输入 $\boldsymbol{z}_{\boldsymbol{U}}$的次数为 ℓ 的球谐函数之间的最优相关性，

$$\|\mathbf{\Gamma}_{\boldsymbol{U},\ell}\|_{\mathrm{op}} = \sup_{T\in L^2(\nu^Y_{d,\boldsymbol{U}}),\ \psi\in\mathscr{S}_{d_{\boldsymbol{U}},\ell}} \frac{\mathbb{E}_{\mathbb{P}_{\nu_d}}[T(\boldsymbol{y}_{\boldsymbol{U}})\psi(\boldsymbol{z}_{\boldsymbol{U}})]^2}{\|T\|^2_{L^2}\|\psi\|^2_{L^2}}.$$

例如，在 [JMS24，命题 A1] 中也出现了类似的变分表示，用于学习稀疏函数；而 [DPVLB24，命题 26]则用于学习高斯单指标模型。

定理*1*（*a*）的证明。论证采用标准的二阶矩方法 [JMS24、JKMS25、Mis25]。令 $\mathcal{A}\in\mathsf{SQ}(q,\ \tau^2_{\mathrm{SQ}})$，并将 ϕ_1、……、ϕ_q 记为 $\mathcal{A}$ 收到响应 $v_t=\mathbb{E}_{\mathbb{P}_{\nu_d,\ \emptyset}}[\phi_t]$、$t\in[q]$时发出的查询序列。对于这些响应，查询 ϕ_t 是固定、确定的，并且对于 $g\sim\pi_d$，它们与 $g\cdot\boldsymbol{W}_*$ 独立。由并集界和马尔可夫不等式，

$$\begin{aligned}\mathbb{P}_{g\sim\pi_d}\left(\exists t\in[q],|\mathbb{E}_{\mathbb{P}^{g\cdot\boldsymbol{W}_*}_{\nu_d}}[\phi_t]-v_t|>\tau_{\mathrm{SQ}}\right) &\le \frac{q}{\tau^2_{\mathrm{SQ}}}\cdot\sup_{t\in[q]}\mathrm{Var}_{g\sim\pi_d}\left\{\mathbb{E}_{\mathbb{P}^{g\cdot\boldsymbol{W}_*}_{\nu_d}}[\phi_t]\right\}\\ &\le \frac{q}{\tau^2_{\mathrm{SQ}}}\cdot\sup_{\|\phi\|_{L^2(\mathbb{P}_{\nu_d,\emptyset})}\le 1}\mathrm{Var}_{g\sim\pi_d}\left\{\mathbb{E}_{\mathbb{P}^{g\cdot\boldsymbol{W}_*}_{\nu_d}}[\phi]\right\},\end{aligned}$$

这里用到了 $\|\phi_t\|_\infty\le 1$。因此，对于 $g\sim\pi_d$，以正概率，所有响应都保持τ_{SQ}-一致，只要

$$\frac{q}{\tau^2_{\mathrm{SQ}}}<\left[\sup_{\|\phi\|_{L^2(\mathbb{P}_{\nu_d,\emptyset})}\le 1}\mathrm{Var}_{g\sim\pi_d}\left\{\mathbb{E}_{\mathbb{P}^{g\cdot\boldsymbol{W}_*}_{\nu_d}}[\phi]\right\}\right]^{-1}, \tag{52}$$

在这种情况下，$\mathcal{A}$ 无法完成检测任务。

我们来计算（52）的右侧。定义

$$\Delta_\phi(g):=\mathbb{E}_{\mathbb{P}^{g\cdot\boldsymbol{W}_*}_{\nu_d}}[\phi(\boldsymbol{y},\boldsymbol{z})]-\mathbb{E}_{\mathbb{P}_{\nu_d,\emptyset}}[\phi(\boldsymbol{y},\boldsymbol{z})]=\mathbb{E}_{\mathbb{P}_{\nu_d,\emptyset}}\left[\left(\rho(g)\cdot\frac{\mathrm{d}\mathbb{P}^{\boldsymbol{W}_*}_{\nu_d}}{\mathrm{d}\mathbb{P}_{\nu_d,\emptyset}}(\boldsymbol{y},\boldsymbol{z})-1\right)\phi(\boldsymbol{y},\boldsymbol{z})\right].$$

利用调和分解（24）和等变性（23），

$$\rho(g)\cdot\frac{\mathrm{d}\mathbb{P}_{\nu_d}}{\mathrm{d}\mathbb{P}_{\nu_d,\emptyset}}(\boldsymbol{y},\boldsymbol{z})-1=\sum_{\ell=1}^{\infty}\langle g\cdot\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}),\mathcal{H}_{d,\ell}(\boldsymbol{z})\rangle_{\mathsf{F}},$$

这里我们省略了对 $\boldsymbol{W}_*$ 的依赖（它在此处是任意固定的）。因此

$$\Delta_\phi(g)=\sum_{\ell=1}^{\infty}\mathbb{E}_{\nu^Y_d}\left[\langle g\cdot\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}),\boldsymbol{\alpha}(\boldsymbol{y})\rangle_{\mathsf{F}}\right],\qquad \boldsymbol{\alpha}(\boldsymbol{y}):=\mathbb{E}_{\mathbb{P}_{\nu_d,\emptyset}}\left[\phi(\boldsymbol{y},\boldsymbol{z})\mathcal{H}_{d,\ell}(\boldsymbol{z})\mid\boldsymbol{y}\right].$$

由舒尔正交关系（28），

$$\begin{aligned}\mathbb{E}_{g\sim\pi_d}\left[|\Delta_\phi(g)|^2\right] &= \sum_{\ell=1}^{\infty}\frac{1}{N_{d,\ell}}\mathbb{E}_{\boldsymbol{y}_1,\boldsymbol{y}_2\sim\nu^Y_d}\left[\langle\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}_1),\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}_2)\rangle_{\mathsf{F}}\langle\boldsymbol{\alpha}_\ell(\boldsymbol{y}_1),\boldsymbol{\alpha}_\ell(\boldsymbol{y}_2)\rangle_{\mathsf{F}}\right]\\ &= \sum_{\ell=1}^{\infty}\frac{\|\mathbb{E}_{\boldsymbol{y}\sim\nu^Y_d}[\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y})\otimes\boldsymbol{\alpha}_\ell(\boldsymbol{y})]\|^2_{\mathsf{F}}}{N_{d,\ell}}.\end{aligned}$$

因此，

$$\sup_{\|\phi\|_{L^2(\mathbb{P}_{\nu_d,\emptyset})}\le 1} \mathrm{Var}_{g\sim\pi_d}\left\{\mathbb{E}_{\mathbb{P}_{\nu_d}^{g\cdot \boldsymbol{W}_*}}[\phi]\right\} = \sup_{\|\phi\|_{L^2(\mathbb{P}_{\nu_d,\emptyset})}\le 1} \mathbb{E}_{g\sim\pi_d}\left[|\Delta_\phi(g)|^2\right]$$
$$= \sup_{\ell\ge 1} \frac{1}{N_{d,\ell}} \left\{ \sup_{\boldsymbol{\alpha}_\ell:\mathcal{Y}\to \mathrm{TSym}_\ell(\mathbb{R}^d)} \frac{\|\mathbb{E}_{\nu_d^Y}[\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y})\otimes \boldsymbol{\alpha}_\ell(\boldsymbol{y})]\|_{\mathsf{F}}^2}{\mathbb{E}_{\nu_d^Y}[\|\boldsymbol{\alpha}_\ell(\boldsymbol{y})\|_{\mathsf{F}}^2]} \right\}.$$

根据标准的表示 表示定理，只需对某个 $\mathcal{L}\in \mathbb{L}_{d,\ell}$ 取 $\boldsymbol{\alpha}(\boldsymbol{y}) = \mathcal{L}[\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y})]$。因此

$$\sup_{\boldsymbol{\alpha}_\ell:\mathcal{Y}\to \mathrm{TSym}_\ell(\mathbb{R}^d)} \frac{\|\mathbb{E}[\boldsymbol{\xi}_{\emptyset,\ell}\otimes \boldsymbol{\alpha}_\ell]\|_{\mathsf{F}}^2}{\mathbb{E}[\|\boldsymbol{\alpha}_\ell\|_{\mathsf{F}}^2]} = \sup_{\mathcal{L}\in\mathbb{L}_{d,\ell}} \frac{\|\mathbb{E}[\boldsymbol{\xi}_{\emptyset,\ell}\otimes \mathcal{L}[\boldsymbol{\xi}_{\emptyset,\ell}]]\|_{\mathsf{F}}^2}{\mathbb{E}[\|\mathcal{L}[\boldsymbol{\xi}_{\emptyset,\ell}]\|_{\mathsf{F}}^2]}$$
$$= \sup_{\mathcal{L}\in\mathbb{L}_{d,\ell}} \frac{\langle \mathcal{L}\mathcal{L}^*, \boldsymbol{\Gamma}_{\emptyset,\ell}\otimes_\ell \boldsymbol{\Gamma}_{\emptyset,\ell}\rangle_{\mathsf{F}}}{\langle \mathcal{L}\mathcal{L}^*, \boldsymbol{\Gamma}_{\emptyset,\ell}\rangle_{\mathsf{F}}} = \|\boldsymbol{\Gamma}_{\emptyset,\ell}\|_{\mathrm{op}},$$

其中我们用到了 $\boldsymbol{\Gamma}_{\emptyset,\ell} = \mathbb{E}[\boldsymbol{\xi}_{\emptyset,\ell}\otimes\boldsymbol{\xi}_{\emptyset,\ell}]$。将其与（52）结合，即得所需的 SQ 下界。 □

定理 *1.(b)* 的证明。该结果可通过改编 [JKMS25 的定理 5] 的证明得到。为方便起见，令 $\mathsf{A}_\star := \mathsf{sAlign}(\nu_d\|\nu_{d,\emptyset})$。将似然比的 D 次投影展开为

$$\mathcal{R}_{\le D}(\{\boldsymbol{y}_i,\boldsymbol{z}_i\}_{i\in[m]}) = \sum_{\ell_1+\ldots+\ell_m\le D} \mathbb{E}_{g\sim\pi_d}\left[\prod_{i\in[m]} \langle \boldsymbol{\xi}_{\emptyset,\ell_i}(\boldsymbol{y}_i), \rho(g)\cdot \mathcal{H}_{d,\ell_i}(\boldsymbol{z}_i)\rangle_{\mathsf{F}}\right],$$

从而

$$\|\mathcal{R}_{\le D}\|_{L^2}^2 = 1 + \sum_{s=1}^{D}\binom{m}{s} \sum_{\substack{1\le \ell_1,\ldots,\ell_s\le D\\ \ell_1+\ldots+\ell_s\le D}} \mathbb{E}_{g\sim\pi_d}\left[\prod_{i\in[s]} F_{\ell_i}(g)\right],$$

其中我们引入了矩阵系数函数

$$F_{\ell_i}(g) := \left\langle \boldsymbol{\Gamma}_{\emptyset,\ell_i}, \mathbb{E}_{\boldsymbol{z}}\left[[\rho(g)\cdot \mathcal{H}_{d,\ell_i}(\boldsymbol{z})]\otimes \mathcal{H}_{d,\ell_i}(\boldsymbol{z})\right]\right\rangle_{\mathsf{F}}.$$

每个 F_ℓ 都属于子空间 $\mathcal{M}_{d,\ \ell}$ 其次数为 ℓ的矩阵系数。根据超收缩性（29）和 Hölder 不等式，¨

$$\mathbb{E}_{g\sim\pi_d}\left[\prod_{i\in[s]} F_{\ell_i}(g)\right] \le \prod_{i\in[s]} \|F_{\ell_i}\|_{L^s(\pi_d)} \le 2^{\frac{s-2}{2}} \prod_{i\in[s]} (s-1)^{\gamma_d(\ell_i)} \|F_{\ell_i}\|_{L^2}.$$

接下来注意到

$$F_\ell(g) = \mathbb{E}_{\boldsymbol{y}\sim\nu_d^Y}\left[\langle g^{-1}\cdot \boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}), \boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y})\rangle_{\mathsf{F}}\right],$$

因此，根据舒尔正交关系（28），

$$\|F_\ell\|_{L^2(\pi_d)}^2 = \mathbb{E}_{\boldsymbol{y}_1,\boldsymbol{y}_2\sim\nu_d^Y}\left[\mathbb{E}_g\left[\langle g\cdot \boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}_1), \boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}_1)\rangle_{\mathsf{F}} \langle g\cdot \boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}_2), \boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}_2)\rangle_{\mathsf{F}}\right]\right]$$
$$= \frac{1}{N_{d,\ell}} \mathbb{E}_{\boldsymbol{y}_1,\boldsymbol{y}_2\sim\nu_d^Y}\left[\langle \boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}_1), \boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}_2)\rangle_{\mathsf{F}}^2\right] = \frac{\|\boldsymbol{\Gamma}_{\emptyset,\ell}\|_{\mathsf{F}}^2}{N_{d,\ell}}.$$

使用 $\gamma_d(\ell_i) \le 4\ell_i$ 得到 $\ell_i \le D \le d-2$，我们有

$$\|\mathcal{R}_{\le D}\|_{L^2}^2 - 1 \le \sum_{s=1}^{D}\binom{m}{s} 2^{s/2} \sum_{\substack{1\le \ell_1,\ldots,\ell_s\le D\\ \ell_1+\ldots+\ell_s\le D}} \prod_{i\in[s]} s^{4\ell_i} \frac{\|\boldsymbol{\Gamma}_{\emptyset,\ell_i}\|_{\mathsf{F}}}{\sqrt{N_{d,\ell_i}}} \le \sum_{s=1}^{D}\binom{m}{s} \rho(s,D)^s,$$

其中

$$\rho(s,D) := \sqrt{2}\sum_{\ell=1}^{D} s^{4\ell}\frac{\|\mathbf{\Gamma}_{\emptyset,\ell}\|_{\mathsf{F}}}{\sqrt{N_{d,\ell}}}.$$

根据假设

$$\frac{\|\mathbf{\Gamma}_{\emptyset,\ell}\|_{\mathsf{F}}}{\sqrt{N_{d,\ell}}} \le \mathsf{A}_\star \le Cd^{p/2}, \qquad \text{for all } \ell > 1. \tag{53}$$

我们还有

$$\|\mathbf{\Gamma}_{\emptyset,\ell}\|_{\mathrm{op}} = \sup_{\boldsymbol{A}\in \mathrm{TSym}_\ell(\mathbb{R}^d)} \frac{\mathbb{E}[\langle \boldsymbol{A}, \boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y})\rangle_{\mathsf{F}}^2]}{\|\boldsymbol{A}\|_{\mathsf{F}}^2} \le \sup_{\boldsymbol{A}\in \mathrm{TSym}_\ell(\mathbb{R}^d)} \frac{\mathbb{E}[\langle \boldsymbol{A}, \mathcal{H}_{d,\ell}(\boldsymbol{z})\rangle_{\mathsf{F}}^2]}{\|\boldsymbol{A}\|_{\mathsf{F}}^2} = 1,$$

其中用到了 $\boldsymbol{\xi}_{\emptyset,\,\ell}(\boldsymbol{y}) = \mathbb{E}[\mathcal{H}_{d,\ell}(\boldsymbol{z})|\boldsymbol{y}]$，Jensen 不等式，以及最后一个等式中的等距映射（22）。根据引理 20，$\boldsymbol{\xi}_{\emptyset,\,\ell}$ 的支撑集所在空间的维数至多为 s^ℓ。因此，$\|\mathbf{\Gamma}_{\emptyset,\,\ell}\|_{\mathsf{F}} \le s^{\ell/2}$，由此得到另一个界

$$\frac{\|\mathbf{\Gamma}_{\emptyset,\ell}\|_{\mathsf{F}}}{\sqrt{N_{d,\ell}}} \le C\frac{s^{\ell/2}d^{p/2}}{\mathsf{A}_\star\sqrt{N_{d,\ell}}}. \tag{54}$$

将求和拆分为 $\ell \le p$ 和 $\ell > p$ 两部分，并使用界（53）和（54），

$$\rho(s,D) \le \sqrt{2}\sum_{\ell=1}^{p}\frac{s^{4\ell}}{\mathsf{A}_\star} + C\sum_{\ell=p+1}^{D}\frac{s^{9\ell/2}d^{p/2}}{\mathsf{A}_\star\sqrt{N_{d,\ell}}} \le \frac{s^{9p/2}}{\mathsf{A}_\star}\left[\sqrt{2}p + \sum_{\ell=p+1}^{D}\frac{D^{9(\ell-p)/2}d^{p/2}}{\sqrt{N_{d,\ell}}}\right].$$

对于 $\ell \le D \le \sqrt{d}$，我们有 $N_{d,\,\ell} \ge c(d/\ell)^\ell$，其中某个常数为 $c > 0$，并且

$$\sum_{\ell=p+1}^{D}\frac{D^{9(\ell-p)/2}d^{p/2}}{\sqrt{N_{d,\ell}}} \le C\sum_{\ell=p+1}^{D}\ell^{p/2}(D^9\ell/d)^{(\ell-p)/2} \le C'\frac{D^{p/2+1}}{\sqrt{d}}.$$

假设 $D = o_d(d^{1/(p+2)})$，我们可推出

$$\rho(s,D) \le 2p\frac{s^{9p/2}}{\mathsf{A}_\star}.$$

因此，

$$\|\mathcal{R}_{\le D}\|_{L^2}^2 - 1 \le \sum_{s=1}^{D}\binom{n}{s}\frac{s^{9sp/2}}{\mathsf{A}_\star^s}(2p)^s \le \sum_{s=1}^{D}\left(2ep\frac{nD^{9p/2-1}}{\mathsf{A}_\star}\right)^s.$$

因此，若 $n = o_d(\mathsf{A}_\star/D^{9p/2-1})$，则 $\|\mathcal{R}_{\le D}\|_{L^2}^2 = 1 + o_d(1)$，证明完毕。 □

B 调和张量展开算法的分析

本节旨在分析第 4 节所述的迭代张量展开算法。我们首先考察张量展开的一个步骤（算法 1），并分别分析对称情形 $a = b$ 和非对称情形 $a \neq b$。我们在附录 3 中证明非对称情形下的定理，在附录 B.1中证明对称情形下的定理。随后，我们在附录 B.2中研究多步过程：通过对一系列约化球面 MIM 迭代应用张量展开得到该过程（算法 B.3），并证明定理 2。最后，我们在附录 4中分析这些算法的运行时间B.4。

在本节中，我们将 c、c'、C、$C' > 0$ 记为仅依赖于 s ℓ 以及假设中的常数的一般常数。特别地，这些常数允许在不同的推导行之间发生变化。我们进一步将 $X \lesssim Y$ 表示为：若存在这样的常数 $C > 0$ 使得 $X \leq CY$。

B.1 张量展开的一步：非对称情形

我们首先考虑采用非对称展开的算法 1 $a \neq b$，即（a，b）$=$（1，0），当 $\ell = 1$时，或$1 \leq a < b$ 当 $a + b = \ell$ 时 $\ell \geq 3$。回顾一下，我们定义了经验展开矩阵

$$\widehat{\boldsymbol{M}} = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} K(\boldsymbol{y}_i, \boldsymbol{y}_j) \mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_i)) \mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_j))^{\mathsf{T}} \in \mathbb{R}^{d^a \times d^a}.$$

我们的证明分为三个步骤。首先，我们控制 $\mathbb{E}[\widehat{\boldsymbol{M}}]$ 的顶部特征向量（引理 1），并证明它们的缩并确实恢复了信号子空间 $\boldsymbol{U}_0 \subseteq \boldsymbol{W}$（引理 2）。然后，我们对 $\|\widehat{\boldsymbol{M}} - \mathbb{E}[\widehat{\boldsymbol{M}}]\|_{\mathrm{op}}$ 进行界定（引理 3）。最后，定理 3 由 Davis-Kahan定理的标准应用得到。

我们首先分析期望 $\mathbb{E}[\widehat{\boldsymbol{M}}]$。令 $\{\mu_j\}_{j \in [D_{\mathsf{s},\ell}]}$ 和 $\{\boldsymbol{V}_j\}_{j \in [D_{\mathsf{s},\ell}]}$ 分别为 $\boldsymbol{\Upsilon}_\emptyset$ 的特征值和特征函数，ℓ ：$= \mathbb{E}[\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y}) \otimes \boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y})]$，其定义见（40）。令 $\mathsf{r} \in [D_{\mathsf{s},\ell}]$ 表示假设3中的秩。定义张量

$$\boldsymbol{Z}_0 = \sum_{j=1}^{\mathsf{r}} (\boldsymbol{W}^{\otimes \ell} \boldsymbol{V}_j) \otimes (\boldsymbol{W}^{\otimes \ell} \boldsymbol{V}_j) \quad \text{and} \quad \boldsymbol{Z}_+ = \sum_{j=\mathsf{r}+1}^{D_{\mathsf{s},\ell}} (\boldsymbol{W}^{\otimes \ell} \boldsymbol{V}_j) \otimes (\boldsymbol{W}^{\otimes \ell} \boldsymbol{V}_j), \tag{55}$$

两者都属于 $(\mathbb{R}^d)^{\otimes 2\ell}$。令 $\boldsymbol{\Pi}_0^{(a)}$ 和 $\boldsymbol{\Pi}_+^{(a)}$ 分别表示 $\mathbb{R}^{d^a}$ 中的正交投影算子，其值域分别为 $\mathrm{Mat}_{a,\ 2\ell-a}(\boldsymbol{Z}_0)$ 和 $\mathrm{Mat}_{a,\ 2\ell-a}(\boldsymbol{Z}_+)$。下一个引理根据投影算子 $\boldsymbol{\Pi}_0^{(a)}$ 和 $\boldsymbol{\Pi}_+^{(a)}$ 控制 $\mathbb{E}[\widehat{\boldsymbol{M}}]$ 的顶部特征子空间。

引理 1. 在假设 *3, 4.(K1)* 和 *4.(K2)*下，存在常数 $c, C, C' > 0$，它们仅依赖于 s、ℓ以及这些假设中的常数，使得对所有 $d \geq C'$ 和谱间隙$\gamma > 1$,

$$c\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2 \boldsymbol{\Pi}_0^{(a)} \preceq \mathbb{E}[\widehat{\boldsymbol{M}}] \preceq C\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2 \left(\boldsymbol{\Pi}_0^{(a)} + \frac{1}{\gamma} \boldsymbol{\Pi}_+^{(a)} + \frac{1}{\sqrt{d}} \mathbf{I}_{d^a} \right).$$

对于足够大的 d，该引理表明，$\mathbb{E}[\widehat{\boldsymbol{M}}]$ 的 $\mathsf{t} := \mathrm{rank}(\mathrm{Mat}_{a,\ 2\ell-a}(\boldsymbol{Z}_0))$ 个顶部特征向量近似张成 $\mathrm{Mat}_{a,\ 2\ell-a}(\boldsymbol{Z}_0)$ 的列空间。为了将这个子空间映射回原始空间 $\mathbb{R}^d$，我们将每个特征向量与自身进行收缩，得到一个 $d \times d$ 矩阵。下一个引理表明，从 $\mathrm{span}(\mathrm{Mat}_{a,\ 2\ell-a}(\boldsymbol{Z}_0))$ 开始执行这一过程，确实能够恢复 $\boldsymbol{U}_0$；$\boldsymbol{U}_0$ 定义为 $\mathrm{Mat}_{1,\ 2\ell-1}(\boldsymbol{Z}_0)$ 的列空间。

引理 2. 记作 $\mathsf{t} = \mathrm{rank}(\mathrm{Mat}_{a,\ 2\ell-a}(\boldsymbol{Z}_0))$。设 $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{\mathsf{t}}] \in \mathbb{R}^{d^a \times \mathsf{t}}$ 是 的列空间的任意一组正交归一基。那么，

$$\mathrm{span}\left(\sum_{i=1}^{\mathsf{t}} \mathrm{Mat}_{1,a-1}(\boldsymbol{v}_i)\mathrm{Mat}_{1,a-1}(\boldsymbol{v}_i)^{\mathsf{T}}\right) = \mathrm{span}(\mathrm{Mat}_{1,2\ell-1}(\boldsymbol{Z}_0)) = \mathrm{span}(\boldsymbol{U}_0).$$

上述两个引理的证明见下文第 B.1.1 节。

$\widehat{\ }$ $\widehat{\ }$引理 *3*（$\boldsymbol{M} - \mathbb{E}\boldsymbol{M}$ 的集中性，条件为 $a \neq b$）。在假设 ***4.(K1)*** 和 ***4.(K3)*** 下，存在常数 c、c'、C、$C' > 0$，它们仅取决于 ℓ、s 以及这些假设中的常数，使得对于任意满足 $d \geq C'$ 且 $d^{\ell/2} \leq n \leq \exp(d^{c'})$ 的 n、d，以至少 $1 - e^{-d^c}$ 的概率，有：

$$\left\|\widehat{\boldsymbol{M}} - \mathbb{E}\left[\widehat{\boldsymbol{M}}\right]\right\|_{\mathrm{op}} \leq C\left[\frac{d^{\ell/2\vee 1}}{n} + \frac{d^{\ell/4\vee 1/2}\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}}{\sqrt{n}}\right].$$

该引理的证明见第 B.1.2 节。注意，稍长一些的论证可以在引理 4 中去除有限秩假设(K3)（从而也可在定理3 中去除该假设3）。不过，上述引理已足以满足本文的需要。

现在我们可以证明非对称情形下的定理 3了。

定理*3* 的证明，其中 $a \neq b$。回忆一下，$\mathsf{t} = \mathrm{rank}(\mathrm{Mat}_{a,2\ell-a}(\boldsymbol{Z}_0))$. 令 $\hat{}\boldsymbol{V} \in \mathbb{R}^{d^a \times \mathsf{t}}$ 为 $\widehat{\boldsymbol{M}}$的前 t 个特征向量，并令 $\boldsymbol{V} \in \mathbb{R}^{d^a \times \mathsf{t}}$ 为 $\mathrm{span}(\boldsymbol{\Pi}_0^{(a)}) = \mathrm{span}(\mathrm{Mat}_{a,2\ell-a}(\boldsymbol{Z}_0))$ 的一组正交归一基。对于任意矩阵 $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{\mathsf{t}}] \in \mathbb{R}^{d^a \times \mathsf{t}}$，引入

$$F(\boldsymbol{A}) := \left[\mathrm{Mat}_{1,a-1}(\boldsymbol{a}_1), \ldots, \mathrm{Mat}_{1,a-1}(\boldsymbol{a}_{\mathsf{t}})\right] \in \mathbb{R}^{d \times (d^{a-1}\mathsf{t})}.$$

根据引理2，$\mathrm{span}(F(\boldsymbol{V})F(\boldsymbol{V})^{\mathsf{T}}) = \mathrm{span}(\boldsymbol{U}_0)$。注意，$F$是一个线性算子，并且相对于弗罗贝尼乌斯范数是等距映射。因此，对于任意正交矩阵 $\boldsymbol{Q} \in \mathbb{R}^{\mathsf{t}\times\mathsf{t}}$，

$$\begin{aligned} F(\boldsymbol{V})F(\boldsymbol{V})^{\mathsf{T}} - F(\hat{\boldsymbol{V}})F(\hat{\boldsymbol{V}})^{\mathsf{T}} &= F(\boldsymbol{V})F(\boldsymbol{V})^{\mathsf{T}} - F(\hat{\boldsymbol{V}}\boldsymbol{Q})F(\hat{\boldsymbol{V}}\boldsymbol{Q})^{\mathsf{T}} \\ &= F(\boldsymbol{V} - \hat{\boldsymbol{V}}\boldsymbol{Q})F(\boldsymbol{V})^{\mathsf{T}} + F(\hat{\boldsymbol{V}}\boldsymbol{Q})F(\boldsymbol{V} - \hat{\boldsymbol{V}}\boldsymbol{Q})^{\mathsf{T}}, \end{aligned}$$

并且由于 $\|\boldsymbol{V}\|_{\mathsf{F}} = \|\hat{}\boldsymbol{V}\|_{\mathsf{F}} = \sqrt{\mathsf{t}}$,

$$\left\|F(\boldsymbol{V})F(\boldsymbol{V})^{\mathsf{T}} - F(\hat{\boldsymbol{V}})F(\hat{\boldsymbol{V}})^{\mathsf{T}}\right\|_{\mathsf{F}} \leq 2\sqrt{\mathsf{t}}\,\|\boldsymbol{V} - \hat{\boldsymbol{V}}\boldsymbol{Q}\|_{\mathsf{F}}.$$

对 $\boldsymbol{Q} \in \mathcal{O}_{\mathsf{t}}$ 取下确界，并利用普罗克鲁斯特距离与投影距离之间的标准关系，

$$\inf_{\boldsymbol{Q}\in\mathcal{O}_{\mathsf{t}}} \|\boldsymbol{V} - \hat{\boldsymbol{V}}\boldsymbol{Q}\|_{\mathsf{F}} \leq \sqrt{2\mathsf{t}}\,\|\boldsymbol{V}\boldsymbol{V}^{\mathsf{T}} - \hat{\boldsymbol{V}}\hat{\boldsymbol{V}}^{\mathsf{T}}\|_{\mathrm{op}},$$

我们得出结论：

$$\|F(\boldsymbol{V})F(\boldsymbol{V})^{\mathsf{T}} - F(\hat{\boldsymbol{V}})F(\hat{\boldsymbol{V}})^{\mathsf{T}}\|_{\mathsf{F}} \leq 2\sqrt{2\,\mathsf{t}}\,\|\boldsymbol{\Pi}_0^{(a)} - \hat{\boldsymbol{V}}\hat{\boldsymbol{V}}^{\mathsf{T}}\|_{\mathrm{op}}.$$

将 $\boldsymbol{\Pi}_0^{(a)} - \hat{}\boldsymbol{V}\hat{}\boldsymbol{V}^{\mathsf{T}} = \boldsymbol{\Delta}_1 + \boldsymbol{\Delta}_2$ 分解，其中

$$\boldsymbol{\Delta}_1 := \boldsymbol{\Pi}_0^{(a)} - \boldsymbol{\Pi}_{\mathrm{pop}}^{(a)}, \qquad \boldsymbol{\Delta}_2 := \boldsymbol{\Pi}_{\mathrm{pop}}^{(a)} - \hat{\boldsymbol{V}}\hat{\boldsymbol{V}}^{\mathsf{T}},$$

且 $\boldsymbol{\Pi}_{\mathrm{pop}}^{(a)}$ 是到 $\mathbb{E}[\widehat{\boldsymbol{M}}]$的前 t 个特征子空间的投影算子。我们可以写作

$$\|\boldsymbol{\Delta}_1\|_{\mathrm{op}} = \|(\mathbf{I} - \boldsymbol{\Pi}_0^{(a)})\boldsymbol{\Pi}_{\mathrm{pop}}^{(a)}\|_{\mathrm{op}}, \qquad \|\boldsymbol{\Delta}_2\|_{\mathrm{op}} = \|(\mathbf{I} - \boldsymbol{\Pi}_{\mathrm{pop}}^{(a)})\hat{\boldsymbol{V}}\|_{\mathrm{op}},$$

我们使用 Davis-Kahan定理对其进行界定（例如，参见 [Bha97中的 Wedin正弦角定理，定理 VII.5。9]）。对于 $\mathbf{\Delta}_1$，我们比较 t 个前特征子空间：分别来自 $\mathbb{E}[\boldsymbol{M}]$ 和

$$\boldsymbol{B} = \mathbf{\Pi}_0^{(a)}\mathbb{E}[\widehat{\boldsymbol{M}}]\mathbf{\Pi}_0^{(a)} + \mathbf{\Pi}_{0,\perp}^{(a)}\mathbb{E}[\widehat{\boldsymbol{M}}]\mathbf{\Pi}_{0,\perp}^{(a)},$$

其中 $\mathbf{\Pi}_0^{(a)}, \perp = \mathbf{I} - \mathbf{\Pi}_0^{(a)}$。根据引理 1，存在 $c,\ C > 0$，使得对于所有 $d,\ \gamma > C,\ \boldsymbol{B}$ 的前 t个特征子空间恰好为 $\mathbf{\Pi}_0^{(a)}$，且前t个特征值与其余特征值之间的间隔至少为 $c\|\boldsymbol{\xi}_{\emptyset,\ \ell}\|_{L^2}^2$。此外，

$$\|\mathbb{E}[\widehat{\boldsymbol{M}}] - \boldsymbol{B}\|_{\mathrm{op}} \le 2\|\mathbf{\Pi}_0^{(a)}\mathbb{E}[\widehat{\boldsymbol{M}}]\mathbf{\Pi}_{0,\perp}^{(a)}\|_{\mathrm{op}} \le 2\|\mathbb{E}[\widehat{\boldsymbol{M}}]\|_{\mathrm{op}}^{1/2}\|\mathbf{\Pi}_{0,\perp}^{(a)}\mathbb{E}[\widehat{\boldsymbol{M}}]\mathbf{\Pi}_{0,\perp}^{(a)}\|_{\mathrm{op}}^{1/2} \le C\frac{\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2}{\sqrt{\gamma}}.$$

由Davis-Kahan定理可得

$$\|\mathbf{\Delta}_1\|_{\mathrm{op}} \le C\frac{\|\mathbb{E}[\widehat{\boldsymbol{M}}] - \boldsymbol{B}\|_{\mathrm{op}}}{\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2} \le C'\frac{1}{\sqrt{\gamma}}. \tag{56}$$

对于 $\mathbf{\Delta}_2$，根据引理 3，存在 $c,\ c',\ C,\ C' > 0$ 使得其成立的概率至少为 $1 - e^{-d^c}$，

$$\|\mathbf{\Delta}_2\|_{\mathrm{op}} \le C\frac{\|\widehat{\boldsymbol{M}} - \mathbb{E}[\widehat{\boldsymbol{M}}]\|_{\mathrm{op}}}{\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2} \le \sqrt{\delta}, \tag{57}$$

其中我们记作 $\delta = C'd^{\ell/2\vee 1}/(n\|\boldsymbol{\xi}_{\emptyset,\ \ell}\|_{L^2}^2)$，并假设 $d \ge C'$, $n \le \exp(d^{c'})$，以及 $\delta \le c'$。

合并（56）和（57）可得

$$\|F(\boldsymbol{V})F(\boldsymbol{V})^{\mathsf{T}} - F(\hat{\boldsymbol{V}})F(\hat{\boldsymbol{V}})^{\mathsf{T}}\|_{\mathsf{F}} \le C\left(\sqrt{\delta} + \frac{1}{\sqrt{\gamma}}\right).$$

将（65）应用于 $a = 1$，可知 $\sum_{i=1}^{\mathsf{t}}\mathrm{Mat}_{1,\ a-1}(\boldsymbol{v}_i)\mathrm{Mat}_{1,\ a-1}(\boldsymbol{v}_i)^{\mathsf{T}}$ 的非零特征值与零之间存在间隔，其下界仅依赖于 s ℓ及假设中的常数。因此，再次应用Davis-Kahan定理，令 $\gamma \ge C$ 和 $\delta \le 1/C$，可得

$$\mathrm{dist}\big(\widehat{\boldsymbol{U}}_0, \boldsymbol{U}_0\big) \le C'\|F(\boldsymbol{V})F(\boldsymbol{V})^{\mathsf{T}} - F(\hat{\boldsymbol{V}})F(\hat{\boldsymbol{V}})^{\mathsf{T}}\|_{\mathsf{F}} \le C''\left(\sqrt{\delta} + \frac{1}{\sqrt{\gamma}}\right),$$

其中我们用到了 $\mathrm{span}(F(\boldsymbol{V})) = \mathrm{span}(\boldsymbol{U}_0)$，且 $\widehat{\boldsymbol{U}}_0$ 是 $F(\widehat{\boldsymbol{V}})F(\widehat{\boldsymbol{V}})^{\mathsf{T}}$ 的前 s_0 个特征向量。 □

B.1.1 引理证明1 和引理 2

在证明引理roving 引理 1之前，我们首先给出以下关于 $\mathbb{E}[\widehat{\boldsymbol{M}}]$ 的低秩近似 ：

引理 4. 在假设 4下 *(K1)*，若 $d \ge 4\ell^4\mathsf{s}$，则

$$\mathbb{E}[\widehat{\boldsymbol{M}}] = (\boldsymbol{W}^{\otimes a})\mathbb{E}\left[K(\boldsymbol{y},\boldsymbol{y}')\mathrm{Mat}_{a,b}\left(\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y})\right)\mathrm{Mat}_{a,b}\left(\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y}')\right)^{\mathsf{T}}\right](\boldsymbol{W}^{\otimes a})^{\mathsf{T}} + \mathbf{\Delta}_{d,\ell}$$

其中 $\boldsymbol{W}^{\otimes a} \in \mathbb{R}^{d^a \times \mathsf{s}^a}$ 是正交 s^a-标架，其定义见 (41) 且

$$\|\mathbf{\Delta}_{d,\ell}\|_{\mathsf{F}} \le B_K\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2\frac{8\ell^2\sqrt{\mathsf{s}}}{\sqrt{d}}.$$

证明。 由于样本 $\{(\boldsymbol{y}_i,\ \boldsymbol{z}_i)\}_{i=1}^n$ 是独立同分布的，并回顾定义 $\boldsymbol{\xi}_{\emptyset,\ \ell}(\boldsymbol{y}) = \mathbb{E}[\mathcal{H}_{d,\ell}(\boldsymbol{z})|\boldsymbol{y}]$,

$$\begin{aligned}\mathbb{E}\big[\widehat{\boldsymbol{M}}\big] &= \mathbb{E}\left[K(\boldsymbol{y},\boldsymbol{y}')\mathrm{Mat}_{a,b}\big(\mathcal{H}_{d,\ell}(\boldsymbol{z})\big)\mathrm{Mat}_{a,b}\big(\mathcal{H}_{d,\ell}(\boldsymbol{z}')\big)^{\mathsf{T}}\right]\\ &= \mathbb{E}\left[K(\boldsymbol{y},\boldsymbol{y}')\mathrm{Mat}_{a,b}\left(\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y})\right)\mathrm{Mat}_{a,b}\left(\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}')\right)^{\mathsf{T}}\right],\end{aligned} \tag{58}$$

其中 $(\boldsymbol{y},\ \boldsymbol{z})$，$(\boldsymbol{y}',\ \boldsymbol{z}') \sim \mathbb{P}_{\nu_d}$ 独立地。利用 $\boldsymbol{\xi}_{\emptyset,\ \ell}(\boldsymbol{y}) = \mathsf{P}_{\mathsf{tf}}(\boldsymbol{W}^{\otimes\ell}\boldsymbol{\zeta}_{\emptyset,\ \ell}(\boldsymbol{y}))$ 以及引理21，我们得到，每当 $d \geq 4\ell^2\mathsf{s}$,

$$\left\|\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}) - \boldsymbol{W}^{\otimes\ell}\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y})\right\|_{\mathsf{F}} \leq \|\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y})\|_{\mathsf{F}}\frac{2\ell^2\sqrt{\mathsf{s}}}{\sqrt{d}} \tag{59}$$

这意味着 $\|\boldsymbol{\xi}_{\emptyset,\ \ell} - \boldsymbol{W}^{\otimes\ell}\boldsymbol{\zeta}_{\emptyset,\ \ell}\|_{L^2} \leq \|\boldsymbol{\xi}_{\emptyset,\ \ell}\|_{L^2}2\ell^2\sqrt{\mathsf{s}/d}$ 并且，根据反三角不等式,

$$\|\boldsymbol{W}^{\otimes\ell}\boldsymbol{\zeta}_{\emptyset,\ell}\|_{L^2} \leq \|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}\left(1 + \frac{2\ell^2\sqrt{\mathsf{s}}}{\sqrt{d}}\right). \tag{60}$$

因此,

$$\mathbb{E}\big[\widehat{\boldsymbol{M}}\big] = \mathbb{E}\left[K(\boldsymbol{y},\boldsymbol{y}')\mathrm{Mat}_{a,b}\big(\boldsymbol{W}^{\otimes\ell}\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y})\big)\mathrm{Mat}_{a,b}\big(\boldsymbol{W}^{\otimes\ell}\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y}')\big)^{\mathsf{T}}\right] + \boldsymbol{\Delta}_{d,\ell},$$

其中由（59）和（60），

$$\|\boldsymbol{\Delta}_{d,\ell}\|_{\mathsf{F}} \leq \|K\|_\infty\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2\frac{4\ell^2\sqrt{\mathsf{s}}}{\sqrt{d}}\left(1 + \frac{2\ell^2\sqrt{\mathsf{s}}}{\sqrt{d}}\right) \leq B_K\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2\frac{8\ell^2\sqrt{\mathsf{s}}}{\sqrt{d}}.$$

只需注意到

$$\mathrm{Mat}_{a,b}\left(\boldsymbol{W}^{\otimes\ell}\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y})\right) = \boldsymbol{W}^{\otimes a}\mathrm{Mat}_{a,b}\left(\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y})\right)(\boldsymbol{W}^{\otimes b})^{\mathsf{T}},$$

以及 $(\boldsymbol{W}^{\otimes b})^{\mathsf{T}}(\boldsymbol{W}^{\otimes b}) = \mathbf{I}_{\mathsf{s}^b}$。至此证明完毕。 □

引理 *1* 的证明。 首先注意到，$\mathbb{E}[\widehat{\boldsymbol{M}}]$ 根据（58）是半正定的。由引理 4，对于任意 $\boldsymbol{v} \in \mathbb{R}^{d^a}$,

$$\boldsymbol{v}^{\mathsf{T}}\mathbb{E}\big[\widehat{\boldsymbol{M}}\big]\boldsymbol{v} = \boldsymbol{v}_{\boldsymbol{W}}^{\mathsf{T}}\mathbb{E}\left[K(\boldsymbol{y},\boldsymbol{y}')\mathrm{Mat}_{a,b}\left(\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y})\right)\mathrm{Mat}_{a,b}\left(\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y}')\right)^{\mathsf{T}}\right]\boldsymbol{v}_{\boldsymbol{W}} + \boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Delta}_{d,\ell}\boldsymbol{v}, \tag{61}$$

其中 $\boldsymbol{v}_{\boldsymbol{W}} := (\boldsymbol{W}^{\otimes a})^{\mathsf{T}}\boldsymbol{v} \in \mathbb{R}^{\mathsf{s}^a}$ 验证了 $\|\boldsymbol{v}_{\boldsymbol{W}}\|_2 \leq \|\boldsymbol{v}\|_2$，利用 $\boldsymbol{W}^{\otimes a} \in \mathbb{R}^{d^a \times \mathsf{s}^a}$ 是一个正交归一的 s^a-标架，并且

$$|\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Delta}_{d,\ell}\boldsymbol{v}| \leq \|\boldsymbol{\Delta}_{d,\ell}\|_{\mathsf{F}}\|\boldsymbol{v}\|_2^2 \leq B_K\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2\frac{8\ell^2\sqrt{\mathsf{s}}}{\sqrt{d}}\|\boldsymbol{v}\|_2^2.$$

令 $\boldsymbol{\zeta}_{\emptyset,\ \ell}(\boldsymbol{y})[\boldsymbol{v}_{\boldsymbol{W}}]$ 表示张量 $\boldsymbol{\zeta}_{\emptyset,\ \ell}(\boldsymbol{y})$ 沿着第一组 s a 个指标与向量 $\boldsymbol{v}_{\boldsymbol{W}} \in \mathbb{R}^{\mathsf{s}^a}$ 的缩并，使得

$$\boldsymbol{v}_{\boldsymbol{W}}^{\mathsf{T}}\mathbb{E}\left[K(\boldsymbol{y},\boldsymbol{y}')\mathrm{Mat}_{a,b}\left(\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y})\right)\mathrm{Mat}_{a,b}\left(\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y}')\right)^{\mathsf{T}}\right]\boldsymbol{v}_{\boldsymbol{W}} = \mathbb{E}[K(\boldsymbol{y},\boldsymbol{y}')\langle\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y})[\boldsymbol{v}_{\boldsymbol{W}}],\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y}')[\boldsymbol{v}_{\boldsymbol{W}}]\rangle_{\mathsf{F}}].$$

根据柯西-施瓦茨不等式和假设 4。(K1),

$$|\mathbb{E}[K(\boldsymbol{y},\boldsymbol{y}')\langle\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y})[\boldsymbol{v}_{\boldsymbol{W}}],\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y}')[\boldsymbol{v}_{\boldsymbol{W}}]\rangle_{\mathsf{F}}]| \leq B_K\|\boldsymbol{\zeta}_{\emptyset,\ell}[\boldsymbol{v}_{\boldsymbol{W}}]\|_{L^2}^2. \tag{62}$$

另一方面，根据假设 4。(K2)，存在 $c_K > 0$ 使得

$$\mathbb{E}[K(\boldsymbol{y},\boldsymbol{y}')\langle\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y})[\boldsymbol{v}_{\boldsymbol{W}}],\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y}')[\boldsymbol{v}_{\boldsymbol{W}}]\rangle_{\mathsf{F}}] \geq c_K\|\boldsymbol{\zeta}_{\emptyset,\ell}[\boldsymbol{v}_{\boldsymbol{W}}]\|_{L^2}^2. \tag{63}$$

它仍然 以刻画 $\mathbb{R}^{\mathsf{s}^a}$ 的子空间，其中量 $\|\boldsymbol{\zeta}_{\emptyset,\,\ell}[\boldsymbol{v}_{\boldsymbol{W}}]\|_{L^2}^2$ 较大 e.

首先，注意我们可以将其分解为

$$\|\boldsymbol{\zeta}_{\emptyset,\ell}[\boldsymbol{v}_{\boldsymbol{W}}]\|_{L^2}^2 = \sum_{j=1}^{D_{\mathsf{s},\ell}} \mu_j \langle \boldsymbol{v}_{\boldsymbol{W}}, \mathrm{Mat}_{a,b}(\boldsymbol{V}_j)\mathrm{Mat}_{a,b}(\boldsymbol{V}_j)^{\mathsf{T}} \boldsymbol{v}_{\boldsymbol{W}} \rangle. \tag{64}$$

令 $\boldsymbol{Z}_0$ 和 $\boldsymbol{Z}_+$ 如式（55）所定义。我们可以写成

$$\boldsymbol{Z}_0 \otimes_{2\ell-a} \boldsymbol{Z}_0 = \sum_{i,j=1}^{\mathsf{r}} \mu_i\mu_j \left((\boldsymbol{W}^{\otimes\ell}\boldsymbol{V}_i) \otimes (\boldsymbol{W}^{\otimes\ell}\boldsymbol{V}_i) \right) \otimes_{2\ell-a} \left((\boldsymbol{W}^{\otimes\ell}\boldsymbol{V}_j) \otimes (\boldsymbol{W}^{\otimes\ell}\boldsymbol{V}_j) \right).$$

由于 $2\ell - a \geq \ell$，上述运算至少对 $\boldsymbol{W}^{\otimes\ell}\boldsymbol{V}_i$ 和 $\boldsymbol{W}^{\otimes\ell}\boldsymbol{V}_j$ 中的 ℓ 个指标进行缩并。鉴于张量 $\{\boldsymbol{W}^{\otimes\ell}\boldsymbol{V}_j\}_{j\in[D_{\mathsf{s},\ell}]}$ 在 $\mathsf{Sym}^\ell(\mathbb{R}^d)$ 中构成一个正交归一集，我们得到

$$\boldsymbol{Z}_0 \otimes_{2\ell-a} \boldsymbol{Z}_0 = \sum_{i=1}^{\mathsf{r}} \mu_i^2 \left((\boldsymbol{W}^{\otimes\ell}\boldsymbol{V}_i) \otimes_{\ell-a} (\boldsymbol{W}^{\otimes\ell}\boldsymbol{V}_i) \right),$$

并且类似地 类似地，对 $\boldsymbol{Z}_+$ 求和 $j = \mathsf{r}+1$，……，$D_{\mathsf{s},\,\ell}$ 即可。引入矩阵 s

$$\boldsymbol{P}_0^{(a)} := \sum_{j=1}^{\mathsf{r}} \mathrm{Mat}_{a,b}(\boldsymbol{V}_j)\mathrm{Mat}_{a,b}(\boldsymbol{V}_j)^{\mathsf{T}}, \qquad \boldsymbol{P}_+^{(a)} := \sum_{j=\mathsf{r}+1}^{D_{\mathsf{s},\ell}} \mathrm{Mat}_{a,b}(\boldsymbol{V}_j)\mathrm{Mat}_{a,b}(\boldsymbol{V}_j)^{\mathsf{T}}.$$

注意，张量 $\{\boldsymbol{V}_i\}_{i\in[D_{\mathsf{s},\ell}]}$ 在 $\mathsf{Sym}^\ell(\mathbb{R}^{\mathsf{s}})$ 中构成一个正交归一集。由紧致性可知，存在常数 $\lambda_{0,1}, \lambda_{0,2}$, $\lambda_{+,1}, \lambda_{+,2} > 0$，它们仅依赖于 $(\mathsf{s}\ \ell, \mathsf{r})$，使得 $\boldsymbol{P}_{(a)\,0}$ 和 $\boldsymbol{P}_+^{(a)}$ 的所有非零特征值分别属于 $[\lambda_{0,1}, \lambda_{0,2}]$ 和 $[\lambda_{+,1}, \lambda_{+,2}]$。因此，我们推得

$$\begin{aligned} \lambda_{0,1}\boldsymbol{\Pi}_0^{(a)} &\preceq \mathrm{Mat}_{a,2\ell-a}(\boldsymbol{Z}_0)\mathrm{Mat}_{a,2\ell-a}(\boldsymbol{Z}_0)^{\mathsf{T}} = (\boldsymbol{W}^{\otimes a})\boldsymbol{P}_0^{(a)}(\boldsymbol{W}^{\otimes a})^{\mathsf{T}} \preceq \lambda_{0,2}\boldsymbol{\Pi}_0^{(a)}, \\ \lambda_{+,1}\boldsymbol{\Pi}_+^{(a)} &\preceq \mathrm{Mat}_{a,2\ell-a}(\boldsymbol{Z}_+)\mathrm{Mat}_{a,2\ell-a}(\boldsymbol{Z}_+)^{\mathsf{T}} = (\boldsymbol{W}^{\otimes a})\boldsymbol{P}_+^{(a)}(\boldsymbol{W}^{\otimes a})^{\mathsf{T}} \preceq \lambda_{+,2}\boldsymbol{\Pi}_+^{(a)}. \end{aligned} \tag{65}$$

由（64），我们推得

$$\begin{aligned} \|\boldsymbol{\zeta}_{\emptyset,\ell}[\boldsymbol{v}_{\boldsymbol{W}}]\|_{L^2}^2 &\leq \mu_1 \langle \boldsymbol{v}, (\boldsymbol{W}^{\otimes a})\boldsymbol{P}_0^{(a)}(\boldsymbol{W}^{\otimes a})^{\mathsf{T}}\boldsymbol{v} \rangle + \mu_{\mathsf{r}+1} \langle \boldsymbol{v}, (\boldsymbol{W}^{\otimes a})\boldsymbol{P}_+^{(a)}(\boldsymbol{W}^{\otimes a})^{\mathsf{T}}\boldsymbol{v} \rangle \\ &\leq \left\langle \boldsymbol{v}, \left[\mu_1\lambda_{0,2}\boldsymbol{\Pi}_0^{(a)} + \frac{c_\mu\mu_1\lambda_{+,2}}{\gamma}\boldsymbol{\Pi}_+^{(a)} \right] \boldsymbol{v} \right\rangle, \end{aligned}$$

and

$$\|\boldsymbol{\zeta}_{\emptyset,\ell}[\boldsymbol{v}_{\boldsymbol{W}}]\|_{L^2}^2 \geq \mu_{\mathsf{r}} \langle \boldsymbol{v}, (\boldsymbol{W}^{\otimes a})\boldsymbol{P}_0^{(a)}(\boldsymbol{W}^{\otimes a})^{\mathsf{T}}\boldsymbol{v} \rangle \geq \lambda_{0,1}c_\mu\mu_1 \langle \boldsymbol{v}, \boldsymbol{\Pi}_0^{(a)}\boldsymbol{v} \rangle,$$

这里用到了假设 3。结合这些界与（62）、（63）和（61），并注意到 $\mu_1 \leq \|\mathbb{E}[\boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y})^{\otimes}\, \boldsymbol{\zeta}_{\emptyset,\ell}(\boldsymbol{y})]\|_{\mathsf{F}} \leq \mathsf{s}^\ell\mu_1$ 与（59）中的关系，从而

$$s^{-\ell}\left(1 - \frac{6\ell^2\sqrt{\mathsf{s}}}{\sqrt{d}}\right)\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2 \leq \mu_1 \leq \left(1 + \frac{6\ell^2\sqrt{\mathsf{s}}}{\sqrt{d}}\right)\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2,$$

即可得到所需结果。 □

Lemma 2 可由张量展开张成空间的一般性质推出 s.

引理5。 设 $\boldsymbol{T} \in (\mathbb{R}^d)^{\otimes \ell}$ 和 $\boldsymbol{X}^{(a)} =$ 矩阵$_{a,\ell-a}(\boldsymbol{T})$ 满足 $2 \le a \le \ell$。记 $\mathsf{t}_a =$ 秩$(\boldsymbol{X}^{(a)})$，并令 $\boldsymbol{Y} = [\boldsymbol{y}_1, \dots, \boldsymbol{y}_{\mathsf{t}_a}] \in \mathbb{R}^{d^a \times \mathsf{t}_a}$ 为张成空间$(\boldsymbol{X}^{(a)})$的任意正交归一基。对于 $1 \le c < a$,

$$\operatorname{span}\left(\sum_{j=1}^{\mathsf{t}_a} \operatorname{Mat}_{c,a-c}(\boldsymbol{y}_j)\operatorname{Mat}_{c,a-c}(\boldsymbol{y}_j)^{\mathsf{T}}\right) = \operatorname{span}(\boldsymbol{X}^{(c)}).$$

证明。 由于每个矩阵 $\operatorname{Mat}_{c,\ a-c}(\boldsymbol{y}_j)\operatorname{Mat}_{c,\ a-c}(\boldsymbol{y}_j)^{\mathsf{T}}$ 都是半正定的，因此我们有

$$\operatorname{span}\left(\sum_{j=1}^{\mathsf{t}_a} \operatorname{Mat}_{c,a-c}(\boldsymbol{y}_j)\operatorname{Mat}_{c,a-c}(\boldsymbol{y}_j)^{\mathsf{T}}\right) = \sum_{j=1}^{\mathsf{t}_a} \operatorname{span}\left(\operatorname{Mat}_{c,a-c}(\boldsymbol{y}_j)\right),$$

其中右侧表示子空间和。令 $\{\boldsymbol{x}_j\}_{j=1}^{d^{\ell-a}}$ 表示$\boldsymbol{X}^{(a)}$的列。由于 $\{\boldsymbol{y}_j\}_{j=1}^{\mathsf{t}_a}$ 是 $\operatorname{span}(\boldsymbol{X}^{(a)})$ 的一组基，族 $\{\boldsymbol{y}_j\}_{j=1}^{\mathsf{t}_a}$ 和 $\{\boldsymbol{x}_j\}_{j=1}^{d^{\ell-a}}$ 张成相同的子空间。因此,

$$\sum_{j=1}^{\mathsf{t}_a} \operatorname{span}\left(\operatorname{Mat}_{c,a-c}(\boldsymbol{y}_j)\right) = \sum_{j=1}^{d^{\ell-a}} \operatorname{span}\left(\operatorname{Mat}_{c,a-c}(\boldsymbol{x}_j)\right).$$

最后，根据构造,

$$\boldsymbol{X}^{(c)} = \left[\operatorname{Mat}_{c,a-c}(\boldsymbol{x}_1) \mid \cdots \mid \operatorname{Mat}_{c,a-c}(\boldsymbol{x}_{d^{\ell-a}})\right],$$

重新排列列的顺序后，这意味着

$$\sum_{j=1}^{d^{\ell-a}} \operatorname{span}\left(\operatorname{Mat}_{c,a-c}(\boldsymbol{x}_j)\right) = \operatorname{span}(\boldsymbol{X}^{(c)}).$$ □

B.1.2 引理3的证明

在假设 4下(K3)，存在 $\mathsf{m} \in \mathbb{N}$ 使得我们可以分解

$$\widehat{\boldsymbol{M}} = \sum_{k=1}^{\mathsf{m}} \widehat{\boldsymbol{M}}_k,$$

其中

$$\widehat{\boldsymbol{M}}_k := \frac{1}{n(n-1)} \sum_{1 \le i \ne j \le n} \mathcal{T}_k(\boldsymbol{y}_i)\mathcal{T}_k(\boldsymbol{y}_j)\operatorname{Mat}_{a,b}\big(\mathcal{H}_{d,\ell}(\boldsymbol{z}_i)\big)\operatorname{Mat}_{a,b}\big(\mathcal{H}_{d,\ell}(\boldsymbol{z}_j)\big)^{\mathsf{T}}.$$

此外，由于根据假设 4.(K1)，对所有 $y \in \mathcal{Y}$ 都有 $\sum_{k=1}^{\mathsf{m}} \mathcal{T}_k(\boldsymbol{y})^2 = K(\boldsymbol{y},\boldsymbol{y}) \le B_K$，因此可得 $\|\mathcal{T}_k\|_\infty \le B_K^{1/2}$。通过对 $k \in [\mathsf{m}]$，使用并集界，只需对满足 $\|\mathcal{T}\|_\infty \le B_K^{1/2}$ 的秩一核 $K(\boldsymbol{y},\boldsymbol{y}') = \mathcal{T}(\boldsymbol{y})\mathcal{T}(\boldsymbol{y}')$ 证明引理 3 即可。注意，可以修改该论证，从而直接对任意有界核（无需有限秩假设）证明引理 3。为简化起见，我们略去这一更为冗长的论证。

下面，我们证明秩一核的引理 3，并将分析拆分为 $1 \le a < b$ 和 $(a,b) = (1,0)$。

引理 3 的证明：情形 $1 \le a < b$。**步骤 1：解耦**。利用 dlPMS 提出的一个如今已成为标准的解耦论证，我们可以将对 [dlPMS99], 的上界估计问题转化为对 $\|\boldsymbol{M} - \mathbb{E}[\boldsymbol{M}]\|_{\mathrm{op}}$ 一个解耦类比量 $\|\boldsymbol{M} - \mathbb{E}[\boldsymbol{M}]\|_{\mathrm{op}}$ 的上界估计问题，其定义如下：

$$\widetilde{\boldsymbol{M}} = \frac{1}{n(n-1)} \sum_{1 \le i \ne j \le n} \mathcal{T}(y_i)\mathcal{T}(\tilde{y}_j)\mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_i))\mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\tilde{\boldsymbol{z}}_j))^{\mathsf{T}} =: \sum_{i=1}^{n} \boldsymbol{A}_i(\tilde{\boldsymbol{B}}_i)^{\mathsf{T}},$$

其中 $\mathcal{D} := \{(\tilde{\boldsymbol{y}}_i,\ \tilde{\boldsymbol{z}}_i)\}_{i=1}^n$ 是 $\mathcal{D} := \{(\boldsymbol{y}_i,\ \boldsymbol{z}_i)\}_{i=1}^n$ 的独立同分布副本，并且

$$\boldsymbol{A}_i = \frac{1}{n}\mathcal{T}(\boldsymbol{y}_i)\mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_i)), \quad \text{and} \quad \tilde{\boldsymbol{B}}_i = \frac{1}{n-1}\sum_{j \ne i} \mathcal{T}(\tilde{\boldsymbol{y}}_j)\mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\tilde{\boldsymbol{z}}_j)).$$

令 $\boldsymbol{E} = \mathbb{E}[\mathcal{T}(\boldsymbol{y})\mathrm{Mat}_{a,\ b}(\mathcal{H}_{d,\ \ell}(\boldsymbol{z}))]$，并注意到 $\mathbb{E}[\widetilde{\boldsymbol{M}}] = \boldsymbol{E}\boldsymbol{E}^{\mathsf{T}}$。我们将其分解 e

$$\widetilde{\boldsymbol{M}} - \boldsymbol{E}\boldsymbol{E}^{\mathsf{T}} = \boldsymbol{\Delta}_1 + \boldsymbol{\Delta}_2 + \boldsymbol{\Delta}_3, \tag{66}$$

其中

$$\boldsymbol{\Delta}_1 := \sum_{i=1}^{n}(\boldsymbol{A}_i - \boldsymbol{E})(\tilde{\boldsymbol{B}}_i - \boldsymbol{E})^{\mathsf{T}}, \qquad \boldsymbol{\Delta}_2 := \sum_{i=1}^{n}(\boldsymbol{A}_i - \boldsymbol{E})\boldsymbol{E}^{\mathsf{T}}, \qquad \boldsymbol{\Delta}_3 := \sum_{i=1}^{n}\boldsymbol{E}(\tilde{\boldsymbol{B}}_i - \boldsymbol{E})^{\mathsf{T}}.$$

注意，$\boldsymbol{\Delta}_1$，$\boldsymbol{\Delta}_2$ 和 $\boldsymbol{\Delta}_3$ 都是独立中心化随机矩阵之和。我们将对这些矩阵以及引理 24 应用，以证明 $\boldsymbol{B}_i - \boldsymbol{E}$的算子范数集中性，其中（66）。

步骤 2：集中性 的集中性 $\boldsymbol{B}_i - \boldsymbol{E}$。我们首先界定（99）中定义的参数。我们有

$$\begin{aligned}\sigma_*(\tilde{\boldsymbol{B}}_i - \boldsymbol{E})^2 &= \sup_{\|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = 1} \mathbb{E}\left[\langle \boldsymbol{u}, (\tilde{\boldsymbol{B}}_i - \boldsymbol{E})\boldsymbol{v}\rangle^2\right] \\ &\le \frac{1}{(n-1)^2}\sum_{j \in [n]\setminus\{i\}} \sup_{\|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = 1} \mathbb{E}\left[\langle \boldsymbol{u}, \mathcal{T}(\boldsymbol{y}_j)\mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_j))\boldsymbol{v}\rangle^2\right].\end{aligned}$$

对于每个 $j \in [n]\setminus\{i\}$ 以及任意 $\boldsymbol{X} \in \mathbb{R}^{d^a \times d^b}$，由假设 4.(K1)、引理 19 以及 $\mathsf{Mat}_{a,b}$ 关于弗罗贝尼乌斯范数是等距映射这一事实可得

$$\mathbb{E}\left[\langle \boldsymbol{X}, \mathcal{T}(\boldsymbol{y}_j)\mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_j))\rangle_{\mathsf{F}}^2\right] \le B_K \left\|\mathsf{P}_{\mathsf{tf}}\left(\mathrm{Mat}_{a,b}^{-1}(\boldsymbol{X})\right)\right\|_{\mathsf{F}}^2 \le B_K\|\boldsymbol{X}\|_{\mathsf{F}}^2. \tag{67}$$

因此，将 67 式应用于 $\boldsymbol{X} = \boldsymbol{u}\boldsymbol{v}^{\mathsf{T}}$，可得 $\sigma_*(\tilde{\boldsymbol{B}}_i - \boldsymbol{E})^2 \le \frac{B_K}{n-1}$ 并且由 100 式可得，$\sigma(\tilde{\boldsymbol{B}}_i - \boldsymbol{E})^2 \le \frac{B_K d^b}{n-1}$。此外，由 67 式可得，

$$v(\tilde{\boldsymbol{B}}_i - \boldsymbol{E})^2 \le \frac{1}{n-1}\sup_{\|\boldsymbol{X}\|_{\mathsf{F}} \le 1} \mathbb{E}\left[\langle \boldsymbol{X}, \mathcal{T}(\boldsymbol{y}_j)\mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_j))\rangle_{\mathsf{F}}^2\right] \le \frac{B_K}{n-1}.$$

此外，注意到这些加项 $\mathcal{T}(\boldsymbol{y}_j)\mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_j))/(n-1)$的算子范数具有确定性上界$\sqrt{B_K N_{d,\ell}}/(n-1)$，这得益于假设4。(K1) 以及式（20）中调和张量的定义（见式（93））。因此，我们直接得到

$$\bar{R}(\tilde{\boldsymbol{B}}_i - \boldsymbol{E})^2 = \mathbb{E}\left[\max_{j \in [n]\setminus\{i\}} \left\|\frac{1}{n-1}\mathcal{T}(\boldsymbol{y}_j)\mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_j)) - \boldsymbol{E}\right\|_{\mathrm{op}}^2\right]$$

$$\leq \frac{4}{(n-1)^2}\mathbb{E}\left[\max_{j\in[n]\setminus\{i\}} \|\mathcal{T}_k(\boldsymbol{y}_j)\mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_j))\|_{\mathrm{op}}^2\right] \leq \frac{4B_k N_{d,\ell}}{(n-1)^2}.$$

应用引理24 及上述界，可得存在一个常数 $C>0$ ，使得对所有 $t\geq 0$ 和 $C\leq d\leq n^{2/\ell}$,

$$\|\tilde{\boldsymbol{B}}_i-\boldsymbol{E}\|_{\mathrm{op}} \leq C\sqrt{B_K}\sqrt{\frac{d^b}{n}}\left(1+\frac{t^{1/2}}{d^{b/2}}+\left(\frac{d^a}{n}\right)^{1/12}t^{2/3}+\left(\frac{d^a}{n}\right)^{1/4}t\right),$$

以至少 $1-(d^a+d^b+1)e^{-t}$ 的概率。选取 $0<c<3(b-a)/48$ 并设定 $t=d^c$，使得对所有 $C\leq d\leq n^{2/\ell}$，以至少 $1-(d^a+d^b+1)e^{-d^c}$ 的概率有 $\|\boldsymbol{B}_i-\boldsymbol{E}\|_{\mathrm{op}}\leq C'\sqrt{B_K d^b/n}$。由于该界对 $i\in[n]$ 一致成立，利用并集界可得

$$\max_{i\in[n]}\|\tilde{\boldsymbol{B}}_i-\boldsymbol{E}\|_{\mathrm{op}} \leq C'\sqrt{B_K}\sqrt{\frac{d^b}{n}}, \tag{68}$$

对所有 $C\leq d\leq n^{2/\ell}$，以至少 $1-n(d^a+d^b+1)e^{-d^c}$ 的概率成立。

步骤 3：的集中性 $\boldsymbol{\Delta}_1$。 令 $\boldsymbol{C}_i := (\boldsymbol{A}_i-\boldsymbol{E})(\tilde{\boldsymbol{B}}_i-\boldsymbol{E})^{\mathsf{T}}$ 表示 $\boldsymbol{\Delta}_1$ 中 $i\in[n]$的各项。我们对 $\boldsymbol{\Delta}_1$ 在给定 $\mathcal{D}$的条件下应用引理24 ，并将讨论限制在 $\mathcal{D}$-可测事件（68）上，该事件发生的概率至少为 $1-n(d^a+d^b+1)e^{-d^c}$。由（67），对每个 $i\in[n]$,

$$\begin{aligned}\mathbb{E}\left[\langle \boldsymbol{X},\boldsymbol{C}_i\rangle_{\mathsf{F}}^2 \mid \tilde{\mathcal{D}}\right] &\leq \frac{1}{n^2}\mathbb{E}\left[\langle \boldsymbol{X}(\tilde{\boldsymbol{B}}_i-\boldsymbol{E}), \mathcal{T}(\boldsymbol{y}_i)\mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_i))\rangle_{\mathsf{F}}^2 \mid \tilde{\mathcal{D}}\right]\\ &\leq \frac{B_K^2}{n^2}\|\boldsymbol{X}(\tilde{\boldsymbol{B}}_i-\boldsymbol{E})\|_{\mathsf{F}}^2 \leq C^2\frac{B_K^2 d^b}{n^3}\|\boldsymbol{X}\|_{\mathsf{F}}^2,\end{aligned}$$

其中我们用到了 $\|\boldsymbol{M}\boldsymbol{X}\|_{\mathsf{F}}\leq\|\boldsymbol{M}\|_{\mathrm{op}}\|\boldsymbol{X}\|_{\mathsf{F}}$ 和（68）。因此，若记 $\sigma_*(\cdot\mid\mathcal{D})$、$v(\cdot\mid\mathcal{D})$ 以及 $R(\cdot\mid\mathcal{D})$ 为将（99）中定义的参数替换为给定 $\mathcal{D}$条件期望后得到的参数，则通过与前面类似的计算可得

$$\sigma_*(\boldsymbol{\Delta}_1\mid\tilde{\mathcal{D}})^2 \leq C^2\frac{B_K^2 d^b}{n^2},\quad \sigma(\boldsymbol{\Delta}_1\mid\tilde{\mathcal{D}})^2 \leq C^2\frac{B_K^2 d^\ell}{n^2},\quad \text{and}\quad v(\boldsymbol{\Delta}_1\mid\tilde{\mathcal{D}})^2 \leq C^2\frac{B_K^2 d^b}{n^2}.$$

为了对各求和项的算子范数进行界定，令 $p\geq 1$ 为一个整数，并注意到根据L^p-范数的单调性和闵可夫斯基不等式,

$$\left\|\max_{i\in[n]}\|\boldsymbol{C}_i\|_{\mathrm{op}}\right\|_{L^{2p}(|\tilde{\mathcal{D}})}^2 \leq \sum_{\alpha,\beta=1}^{d^a}\left\|\max_{i\in[n]}|\langle \boldsymbol{e}_\alpha,\boldsymbol{C}_i\boldsymbol{e}_\beta\rangle|\right\|_{L^{2p}(|\tilde{\mathcal{D}})}^2$$

其中 $\{\boldsymbol{e}_\alpha\}_{\alpha=1}^{d^a}$ 是 $\mathbb{R}^{d^a}$ 的典范基。由于对于任意随机变量 $\{X_i\}_{i=1}^n$，都有 $\max_{i\in[n]}|X_i|\leq(\sum_{i=1}^n|X_i|^{2p})^{1/(2p)}$，

$$\left\|\max_{i\in[n]}\|\boldsymbol{C}_i\|_{\mathrm{op}}\right\|_{L^{2p}(|\tilde{\mathcal{D}})}^2 \leq \sum_{\alpha,\beta=1}^{d^a}\left(\sum_{i=1}^n\mathbb{E}[|\langle \boldsymbol{e}_\alpha,\boldsymbol{C}_i\boldsymbol{e}_\beta\rangle|^{2p}\mid\tilde{\mathcal{D}}]\right)^{1/p}.$$

根据假设4，(K1) 以及超收缩性（30），对于每个 $i\in[n]$ 以及任意 $\alpha,\ \beta\in[d^a]$,

$$\mathbb{E}[|\langle \boldsymbol{e}_\alpha,\boldsymbol{C}_i\boldsymbol{e}_\beta\rangle|^{2p}\mid\tilde{\mathcal{D}}] \leq \frac{B_K^p}{n^{2p}}\mathbb{E}\left[\left|\langle \boldsymbol{e}_\alpha,\mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_i))(\tilde{\boldsymbol{B}}_i-\boldsymbol{E})^{\mathsf{T}}\boldsymbol{e}_\beta\rangle\right|^{2p}\mid\tilde{\mathcal{D}}\right]$$

$$\leq (2p-1)^{\ell p}\frac{B_K^p}{n^{2p}}\mathbb{E}\left[\left|\langle \boldsymbol{e}_\alpha, \mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_i))(\tilde{\boldsymbol{B}}_i-\boldsymbol{E})^\mathsf{T}\boldsymbol{e}_\beta\rangle\right|^2 \mid \tilde{\mathcal{D}}\right]^p.$$

特别地，根据（67），

$$\mathbb{E}[|\langle \boldsymbol{e}_\alpha, \boldsymbol{C}_i\boldsymbol{e}_\beta\rangle|^{2p}\mid\tilde{\mathcal{D}}]\leq (2p-1)^{\ell p}\frac{B_K^{2p}C^{2p}d^{pb}}{n^{3p}},$$

因此，

$$\left\|\max_{i\in[n]}\|\boldsymbol{C}_i\|_{\mathrm{op}}\right\|_{L^{2p}(|\tilde{\mathcal{D}})}\leq\frac{CB_K(2p-1)^{\ell/2}d^{a/2}d^{\ell/2}}{n^{3/2-1/(2p)}}.$$

设 $c\in(0,1-2a/\ell)$ 和 $p=\lfloor d^c\rfloor$。那么，

$$\left\|\max_{i\in[n]}\|\boldsymbol{C}_i\|_{\mathrm{op}}\right\|_{L^{2p}(|\tilde{\mathcal{D}})}\leq\frac{CB_K2^{\ell/2}d^{3\ell/4}}{d^\epsilon n^{3/2}}\exp\left(\frac{\log(n)}{2\lfloor d^c\rfloor}\right)$$

其中 $\epsilon=((1-2c)\ell/2-a)/2>0$。$L^p$ 范数关于 p 的单调性给出

$$\bar{R}(\boldsymbol{\Delta}_1\mid\tilde{\mathcal{D}})\leq\left\|\max_{i\in[n]}\|\boldsymbol{C}_i\|_{\mathrm{op}}\right\|_{L^{2p}(|\tilde{\mathcal{D}})}\leq\frac{CB_K2^{\ell/2}d^{3\ell/4}}{d^\epsilon n^{3/2}}\exp\left(\frac{\log(n)}{2\lfloor d^c\rfloor}\right)$$

而马尔可夫不等式给出

$$\mathbb{P}\left(\max_{i\in[n]}\|\boldsymbol{C}_i\|_{\mathrm{op}}\geq\frac{B_Kd^{3\ell/4}}{d^{\epsilon/2}n^{3/2}}\mid\tilde{\mathcal{D}}\right)\leq\left(\frac{C2^{\ell/2}}{d^{\epsilon/2}}\exp\left(\frac{\log(n)}{2\lfloor d^c\rfloor}\right)\right)^{2\lfloor d^c\rfloor}.$$

假设 $d^{\ell/2}\leq n\leq e^{d^{c'}}$ 对于一个足够小的常数 $c'>0$。条件地在 $\tilde{}$上 $\mathcal{D}$ 应用引理 24，并使用 (68)，可知存在一个常数 $C'>0$ 使得对所有 $d\geq C'$ 和 $t\geq 0$，

$$\|\boldsymbol{\Delta}_1\|_{\mathrm{op}}\leq\frac{C'B_Kd^{\ell/2}}{n}\left(1+\frac{\log(d^b)^{3/4}}{d^{a/4}}+\frac{t^{1/2}}{d^{a/2}}+\frac{t^{2/3}}{d^{\epsilon/12}}+\frac{t}{d^{\epsilon/4}}\right),$$

其概率至少为

$$1-\left(\frac{CB_K^{1/2}2^{\ell/2}}{d^{\epsilon/2}}\exp\left(\frac{\log(n)}{2\lfloor d^c\rfloor}\right)\right)^{2\lfloor d^c\rfloor}-(1+2d^a)e^{-t}-n(d^a+d^b+1)e^{-d^c}.$$

特别地，我们可以取足够小的 $c''>0$，并取 $t=d^{c''}$ 使得

$$\|\boldsymbol{\Delta}_1\|_{\mathrm{op}}\leq\frac{C''B_Kd^{\ell/2}}{n},\tag{69}$$

其概率至少为 $1-e^{-d^{c''}}$。

步骤 4：$\boldsymbol{\Delta}_2$ 和 $\boldsymbol{\Delta}_3$ 的集中性。注意

$$\boldsymbol{E}=\mathbb{E}[\mathcal{T}(\boldsymbol{y})\mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}))]=\mathbb{E}[\mathcal{T}(\boldsymbol{y})\mathrm{Mat}_{a,b}(\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}))],$$

这样我们就可以使用假设 4。(K1)来得到 $\|\boldsymbol{E}\|_{\mathsf{F}}\leq B_K^{1/2}\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}$。采用与推导出 (69) 类似的论证，将 $\max_{i\in[n]}\|\boldsymbol{B}_i-\boldsymbol{E}\|_{\mathrm{op}}$ 的上界替换为上文给出的 $\|\boldsymbol{E}\|_{\mathsf{F}}$ 上界，可知存在常数 C，c，$c'>0$ 使得对于 $d\geq C$ 以及 $d^{\ell/2}\leq n\leq e^{d^{c'}}$，

$$\|\boldsymbol{\Delta}_2\|_{\mathrm{op}},\|\boldsymbol{\Delta}_3\|_{\mathrm{op}}\leq\frac{CB_Kd^{\ell/4}\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}}{\sqrt{n}},\tag{70}$$

其成立的概率至少为 $1-e^{-d^c}$。这些重复的细节从略。

步骤 5：收尾。令 $t \geq 0$。根据 [dlPMS99], 的解耦不等式，存在一个通用常数 $C>0$ 使得

$$\mathbb{P}\left(\|\widehat{\boldsymbol{M}}-\mathbb{E}[\widehat{\boldsymbol{M}}]\|_{\mathrm{op}} \geq t\right) \leq C\mathbb{P}\left(\|\widetilde{\boldsymbol{M}}-\boldsymbol{E}\boldsymbol{E}^{\mathsf{T}}\|_{\mathrm{op}} \geq \frac{t}{C}\right) \leq C\sum_{j\in[3]}\mathbb{P}\left(\|\boldsymbol{\Delta}_j\|_{\mathrm{op}} \geq \frac{t}{3C}\right).$$

结合（69）和（70），可知存在常数 C'，c，$c'>0$ 仅依赖于 s ℓ和假设中的常数，使得对所有 $d \geq C'$ 和 $d^{\ell/2} \leq n \leq e^{d^{c'}}$，

$$\|\widehat{\boldsymbol{M}}-\mathbb{E}[\widehat{\boldsymbol{M}}]\|_{\mathrm{op}} \leq C'\left(\frac{d^{\ell/2}}{n}+\frac{d^{\ell/4}\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}}{\sqrt{n}}\right),$$

以至少 $1-e^{-d^c}$ 的概率成立。证明至此完成。 □

引理 3的证明：情形 $(a,\ b)=(1,\ 0)$。对于 $\ell=1$ (向量空间)，我们只需有 $\mathcal{H}_{d,\ 1}(\boldsymbol{z})=\sqrt{d}\kappa_{d,\ 1}\boldsymbol{z}$。使用与情形 $1 \leq a<b$ 相同的解耦论证，我们引入

$$\widetilde{\boldsymbol{M}}=\frac{d\kappa_{d,1}^2}{n(n-1)}\sum_{1\leq i\neq j\leq n}\mathcal{T}(\boldsymbol{y}_i)\mathcal{T}(\tilde{\boldsymbol{y}}_j)\boldsymbol{z}_i\tilde{\boldsymbol{z}}_j^{\mathsf{T}}=:\sum_{i=1}^n\boldsymbol{a}_i\tilde{\boldsymbol{b}}_i^{\mathsf{T}},$$

˜其中 $\mathcal{D}:=\{(\tilde{\boldsymbol{y}}_i,\ \tilde{\boldsymbol{z}}_i)\}_{i=1}^n$是 $\mathcal{D}:=\{(\boldsymbol{y}_i,\ \boldsymbol{z}_i)\}_{i=1}^n$的独立同分布副本，并且

$$\boldsymbol{a}_i=\frac{\sqrt{d}\kappa_{d,1}}{n}\mathcal{T}(\boldsymbol{y}_i)\boldsymbol{z}_i,\qquad \tilde{\boldsymbol{b}}_i=\frac{\sqrt{d}\kappa_{d,1}}{n-1}\sum_{j\neq i}\mathcal{T}(\tilde{\boldsymbol{y}}_j)\tilde{\boldsymbol{z}}_j.$$

令 $\boldsymbol{e}:=\sqrt{d}\kappa_{d,\ 1}\mathbb{E}[\mathcal{T}(\boldsymbol{y})\boldsymbol{z}]$ 并按照前一种情况$1 \leq a<b$对 $\widetilde{\boldsymbol{M}}-\boldsymbol{e}\boldsymbol{e}^{\mathsf{T}}=\boldsymbol{\Delta}_1+\boldsymbol{\Delta}_2+\boldsymbol{\Delta}_3$ 进行分解，其中

$$\boldsymbol{\Delta}_1:=\sum_{i=1}^n(\boldsymbol{a}_i-\boldsymbol{e})(\tilde{\boldsymbol{b}}_i-\boldsymbol{e})^{\mathsf{T}},\qquad \boldsymbol{\Delta}_2:=\sum_{i=1}^n(\boldsymbol{a}_i-\boldsymbol{e})\boldsymbol{e}^{\mathsf{T}},\qquad \boldsymbol{\Delta}_3:=\sum_{i=1}^n\boldsymbol{e}(\tilde{\boldsymbol{b}}_i-\boldsymbol{e})^{\mathsf{T}}.$$

令 $\boldsymbol{u},\ \boldsymbol{v}\in\mathbb{S}^{d-1}$为任意值。根据假设 4.(K1) 以及 $\|\langle\boldsymbol{z},\ \boldsymbol{u}\rangle\|_{\psi_2}\lesssim 1/\sqrt{d}$这一事实，存在一个通用常数 $C_1>0$，使得 $\|\langle\boldsymbol{a}_i,\ \boldsymbol{u}\rangle\|_{\psi_2}\leq C_1\kappa_{d,\ 1}B_K^{1/2}/n$。在给定 $\tilde{\mathcal{D}}$的条件下，$\boldsymbol{u}^{\mathsf{T}}\boldsymbol{\Delta}_1\boldsymbol{v}$ 是独立的中心化次高斯随机变量之和。在给定 $\mathcal{D}$的条件下应用次高斯霍夫丁不等式，存在一个常数 $C_2>0$，使得

$$\mathbb{P}\left(|\boldsymbol{u}^{\mathsf{T}}\boldsymbol{\Delta}_1\boldsymbol{v}|\geq t\mid\tilde{\mathcal{D}}\right)\leq 2\exp\left(-\frac{C_2nt^2}{\kappa_{d,1}^2B_K\max_{i\in[n]}\|\boldsymbol{b}_i-\boldsymbol{e}\|_2^2}\right),$$

对于任意 $t\geq 0$。另一方面，再次应用霍夫丁不等式可得 $\|\langle\boldsymbol{b}_i,\ \boldsymbol{v}\rangle\|_{\psi_2}\leq C\kappa_{d,\ 1}B_K^{1/2}/\sqrt{n-1}$。利用 $\epsilon=1/4$ 的标准 ϵ-网论证可得

$$\mathbb{P}\left(\|\boldsymbol{b}_i-\boldsymbol{e}\|_2\geq\frac{C_3B_K^{1/2}\kappa_{d,1}\sqrt{d}}{\sqrt{n-1}}\right)\leq 2e^{-d},$$

对于某个充分大的 $C_3>0$。取并集界可得

$$\mathbb{P}\left(|\boldsymbol{u}^{\mathsf{T}}\boldsymbol{\Delta}_1\boldsymbol{v}|\geq t\right)\leq 2\exp\left(-\frac{C_4n(n-1)t^2}{\kappa_{d,1}^4B_K^2d}\right)+2ne^{-d}$$

对于任意 $t \geq 0$，其中 $C_4 > 0$ 是一个常数。取 $t = C_5 \kappa_{d,\ 1}^2 B_K d/n$ 并使用标准的 ϵ-网论证，可得存在常数 C，c，$c' > 0$ 使得

$$\|\mathbf{\Delta}_1\|_{\mathrm{op}} \leq \frac{C\kappa_{d,1}^2 B_K d}{n},$$

以至少 $1 - e^{-d^c}$ 的概率，对任意 $C \leq d \leq n \leq e^{d^{c'}}$ 均成立。运用类似论证，并使用 $\|\boldsymbol{e}\|_2 \leq B_k^{1/2} \|\boldsymbol{\xi}_{\emptyset,\ 1}\|_{L^2}$，可得

$$\|\mathbf{\Delta}_2\|_{\mathrm{op}} \leq \frac{CB_K\sqrt{d}\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}}{\sqrt{n}}, \qquad \|\mathbf{\Delta}_3\|_{\mathrm{op}} \leq \frac{CB_K\sqrt{d}\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}}{\sqrt{n}},$$

以至少 $1 - e^{-d^c}$ 的概率成立。结论与情形 $1 \leq a < b$ 类似。 □

B.2 张量展开的一步：对称情形

现在考虑算法 1 采用对称展开 $a = b = \ell/2$ (ℓ 偶数) 的情形。此时，经验展开矩阵为

$$\widehat{\boldsymbol{M}} = \frac{1}{n^2} \sum_{1 \leq i,j \leq n} K(\boldsymbol{y}_i, \boldsymbol{y}_j) \mathrm{Mat}_{a,a}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_i)) \mathrm{Mat}_{a,a}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_j))^{\mathsf{T}} \in \mathbb{R}^{d^a \times d^a}.$$

在有限秩假设4。(K3)下，可将该矩阵重写为

$$\widehat{\boldsymbol{M}} = \sum_{r=1}^{\mathsf{m}} \widehat{\boldsymbol{S}}_r \widehat{\boldsymbol{S}}_r^{\mathsf{T}}, \qquad \widehat{\boldsymbol{S}}_r := \frac{1}{n} \sum_{1 \leq i \leq n} \mathcal{T}_r(\boldsymbol{y}_i) \mathrm{Mat}_{a,a}(\mathcal{H}_{d,\ell}(\boldsymbol{z}_i)) \in \mathbb{R}^{d^a \times d^a}.$$

我们沿用非对称情形证明中的类似方法，但现在与总体 $\mathbb{E}[\widehat{\boldsymbol{S}}_r]\mathbb{E}[\widehat{\boldsymbol{S}}_r]^{\mathsf{T}}$ 进行比较，而不是与 $\mathbb{E}[\widehat{\boldsymbol{M}}]^{\mathsf{T}}$ 比较。以下两个引理与引理 1 和 3类似。

引理 6。 在假设 *3*和假设 *4*下，存在常数 c，C，$C' > 0$，它们仅依赖于 s、ℓ以及这些假设中的常数，使得对所有 $d \geq C'$ 和 $\gamma > 1$，

$$c\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2 \mathbf{\Pi}_0^{(a)} \preceq \sum_{r=1}^{\mathsf{m}} \mathbb{E}[\widehat{\boldsymbol{S}}_r]\mathbb{E}[\widehat{\boldsymbol{S}}_r]^{\mathsf{T}} \preceq C\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}^2 \left(\mathbf{\Pi}_0^{(a)} + \frac{1}{\gamma}\mathbf{\Pi}_+^{(a)} + \frac{1}{\sqrt{d}}\mathbf{I}_{d^a} \right).$$

该引理可由引理 1 直接推出，只需注意到

$$\sum_{r=1}^{\mathsf{m}} \mathbb{E}[\widehat{\boldsymbol{S}}_r]\mathbb{E}[\widehat{\boldsymbol{S}}_r]^{\mathsf{T}} = \mathbb{E}\left[K(\boldsymbol{y}, \boldsymbol{y}') \mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z})) \mathrm{Mat}_{a,b}(\mathcal{H}_{d,\ell}(\boldsymbol{z}'))^{\mathsf{T}} \right].$$

^集中性部分的处理方式有所不同。我们利用张量空间 TSym 的半单分解展开 $\boldsymbol{S}_r$ $\ell(\mathbb{R}^d) \otimes \mathrm{TSym}^\ell(\mathbb{R}^d)$（在引理 22中已证明）。主导分量对应于一个样本协方差矩阵，我们利用 Bai-Yin 型不等式（适用于重尾随机向量）对其进行界定。其他分量的维度较低，使用与非对称情形中相同的矩阵集中界（引理 24）进行界定。

引理 7 （集中性：对称情形 $a = b$）。 在假设 *4*。*(K1)* 和 *4*。*(K3)*下，存在常数 c，C，$C' > 0$，它们仅依赖于 ℓ、s以及这些假设中的常数，使得对于任意 n，d，若满足 $d \geq C'$ 和 $n \geq d^{\ell/2}$，则以至少 $1 - e^{-d^c}$ 的概率，

$$\left\| \widehat{\boldsymbol{M}} - \sum_{r=1}^{\mathsf{m}} \mathbb{E}[\widehat{\boldsymbol{S}}_r]\mathbb{E}[\widehat{\boldsymbol{S}}_r]^{\mathsf{T}} \right\|_{\mathrm{op}} \leq C \left[\frac{d^{\ell/2}}{n} + \frac{d^{\ell/4}\|\boldsymbol{\xi}_{\emptyset,\ell}\|_{L^2}}{\sqrt{n}} \right].$$

此引理的证明见附录B.2.1 下文。利用这两个引理，采用对称展开的定理3 的证明与采用非对称展开的证明完全相同。

B.2.1 引理的证明7

引理7的证明。同样，依据并集界，只需证明此引理对$\mathsf{m}=1$成立。下文假设 $K(\boldsymbol{y},\ \boldsymbol{y}')=\mathcal{T}(\boldsymbol{y})\mathcal{T}(\boldsymbol{y}')$ 满足 $\|\mathcal{T}\|_\infty \le B_K^{1/2}$，并记 $\widehat{\boldsymbol{S}}:=\widehat{\boldsymbol{S}}_1$。

步骤1：分解 $\widehat{\boldsymbol{S}}$。设 $\boldsymbol{u}\in\mathbb{S}^{d^a-1}$ 为任意的，并用 $\boldsymbol{U}=\mathrm{Mat}_{a,0}^{-1}(\boldsymbol{u})\in(\mathbb{R}^d)^{\otimes a}$表示其张量表示。根据引理22，我们可以分解

$$b_{a,a,0}^{(d)}\langle\mathcal{H}_{d,\ell}(\boldsymbol{z}),\mathsf{P}_{\mathsf{tf}}(\boldsymbol{U})\otimes\mathsf{P}_{\mathsf{tf}}(\boldsymbol{U})\rangle_{\mathsf{F}}=\langle\mathcal{H}_{d,a}(\boldsymbol{z})\otimes\mathcal{H}_{d,a}(\boldsymbol{z}),\mathsf{P}_{\mathsf{tf}}(\boldsymbol{U})\otimes\mathsf{P}_{\mathsf{tf}}(\boldsymbol{U})\rangle_{\mathsf{F}}$$
$$-\sum_{j=1}^{\ell/2}b_{a,a,j}^{(d)}\langle\mathcal{H}_{d,\ell-2j}(\boldsymbol{z}),\mathsf{P}_{\mathsf{tf}}(\boldsymbol{U})\diamond_j\mathsf{P}_{\mathsf{tf}}(\boldsymbol{U})\rangle_{\mathsf{F}},$$

其中系数 $b_{a\,a\,,\,,\,j}^{(d)}=\Theta_d(1)$在引理22中明确给出。为方便起见，记$\boldsymbol{h}_{d,\ a}(\boldsymbol{z}):=\mathrm{Mat}_{a,\ 0}(\mathcal{H}_{d,\ a}(\boldsymbol{z}))\in\mathbb{R}^{d^a-1}$。则，

$$\|\widehat{\boldsymbol{S}}-\mathbb{E}[\widehat{\boldsymbol{S}}]\|_{\mathrm{op}}\le\frac{1}{b_{a,a,0}^{(d)}}\left\|\frac{1}{n}\sum_{i=1}^n\mathcal{T}(\boldsymbol{y}_i)\boldsymbol{h}_{d,a}(\boldsymbol{z}_i)\boldsymbol{h}_{d,a}(\boldsymbol{z}_i)^{\mathsf{T}}-\mathbb{E}\left[\mathcal{T}(\boldsymbol{y})\boldsymbol{h}_{d,a}(\boldsymbol{z})\boldsymbol{h}_{d,a}(\boldsymbol{z})^{\mathsf{T}}\right]\right\|_{\mathrm{op}}$$
$$+\sum_{j=1}^{\ell/2}\sup_{\boldsymbol{u}\in\mathbb{S}^{d^a-1}}\left|\frac{b_{a,a,j}^{(d)}}{b_{a,a,0}^{(d)}}\left\langle\frac{1}{n}\sum_{i=1}^n\mathcal{T}(\boldsymbol{y}_i)\mathcal{H}_{d,\ell-2j}(\boldsymbol{z}_i)-\mathbb{E}[\mathcal{T}(\boldsymbol{y})\mathcal{H}_{d,\ell-2j}(\boldsymbol{z})],\mathsf{P}_{\mathsf{tf}}(\boldsymbol{U})\diamond_j\mathsf{P}_{\mathsf{tf}}(\boldsymbol{U})\right\rangle_{\mathsf{F}}\right|.$$

将右端第一项记为 $\boldsymbol{\Delta}_0$，并对 $j\in[\ell/2]$将求和中的第 j项记为 $\boldsymbol{\Delta}_j$。对任意 $j\in[\ell/2]$，由缩并积 $\diamond_j$ 的定义可得

$$\begin{aligned}&\left|\frac{b_{a,a,j}^{(d)}}{b_{a,a,0}^{(d)}}\left\langle\frac{1}{n}\sum_{i=1}^n\mathcal{T}(\boldsymbol{y}_i)\mathcal{H}_{d,\ell-2j}(\boldsymbol{z}_i)-\mathbb{E}[\mathcal{T}(\boldsymbol{y})\mathcal{H}_{d,\ell-2j}(\boldsymbol{z})],\mathsf{P}_{\mathsf{tf}}(\boldsymbol{U})\diamond_j\mathsf{P}_{\mathsf{tf}}(\boldsymbol{U})\right\rangle_{\mathsf{F}}\right|\\ \lesssim&\left|\left\langle\frac{1}{n}\sum_{i=1}^n\mathcal{T}(\boldsymbol{y}_i)\mathcal{H}_{d,\ell-2j}(\boldsymbol{z}_i)-\mathbb{E}[\mathcal{T}(\boldsymbol{y})\mathcal{H}_{d,\ell-2j}(\boldsymbol{z})],\mathsf{P}_{\mathsf{tf}}(\boldsymbol{U})\otimes_j\mathsf{P}_{\mathsf{tf}}(\boldsymbol{U})\right\rangle_{\mathsf{F}}\right|\\ =&\left|\left\langle\mathrm{Mat}_{a-j,j}(\mathsf{P}_{\mathsf{tf}}(\boldsymbol{U}))^{\mathsf{T}}\left(\widehat{\boldsymbol{S}}^{(\ell-2j)}-\mathbb{E}[\widehat{\boldsymbol{S}}^{(\ell-2j)}]\right)\mathrm{Mat}_{a-j,j}(\mathsf{P}_{\mathsf{tf}}(\boldsymbol{U}))\right\rangle_{\mathsf{F}}\right|\end{aligned}$$

其中我们定义

$$\widehat{\boldsymbol{S}}^{(\ell-2j)}:=\frac{1}{n}\sum_{i=1}^n\mathcal{T}(\boldsymbol{y}_i)\mathrm{Mat}_{a-j,a-j}(\mathcal{H}_{d,\ell-2j}(\boldsymbol{z}_i))\in\mathbb{R}^{d^{a-j}\times d^{a-j}}.$$

由于 $\|\mathrm{Mat}_{a-j,\ j}(\mathsf{P}_{\mathsf{tf}}(\boldsymbol{U}))\|_{\mathsf{F}}\le\|\boldsymbol{U}\|_{\mathsf{F}}=1$，我们得到对任意 $j\in\{1,\dots,\ell/2\}$，

$$\boldsymbol{\Delta}_j\lesssim\left\|\widehat{\boldsymbol{S}}^{(\ell-2j)}-\mathbb{E}[\widehat{\boldsymbol{S}}^{(\ell-2j)}]\right\|_{\mathrm{op}}.$$

步骤 2：集中性，关于 $\boldsymbol{\Delta}_j$，对于 $j\in\{1,\dots,\ell/2\}$。固定 $j\in\{1,\dots,\ell/2\}$。让我们给出出现在（99）中的参数的界。根据假设 4。（K1）以及（93），

$$\|\mathcal{T}(\boldsymbol{y})\mathrm{Mat}_{a-j,a-j}\left(\mathcal{H}_{d,\ell-2j}(\boldsymbol{z})\right)\|_{\mathsf{F}}\le B_K^{1/2}\sqrt{N_{d,\ell-2j}},$$

因此

$$\bar{R}\left(\widehat{\boldsymbol{S}}^{(\ell-2j)}-\mathbb{E}[\widehat{\boldsymbol{S}}^{(\ell-2j)}]\right)^2\le\frac{4B_KN_{d,\ell-2j}}{n^2}.$$

此外，我们可以结合假设 4。(K1) 与引理 19得到

$$\sigma_*\left(\widehat{\boldsymbol{S}}^{(\ell-2j)}-\mathbb{E}[\widehat{\boldsymbol{S}}^{(\ell-2j)}]\right)^2 \le \frac{1}{n^2}\sum_{i=1}^{n}\sup_{\|\boldsymbol{u}\|_2=\|\boldsymbol{v}\|_2=1}\mathbb{E}\left[\left\langle \mathcal{T}(\boldsymbol{y}_i)\mathcal{H}_{d,\ell-2j}(\boldsymbol{z}_i),\mathrm{Mat}^{-1}_{\ell/2-j,\ell/2-j}(\boldsymbol{u}\boldsymbol{v}^{\mathsf{T}})\right\rangle_{\mathsf{F}}^2\right]$$
$$\le \frac{B_K}{n},$$

并且，通过类似的论证，

$$v\left(\widehat{\boldsymbol{S}}^{(\ell-2j)}-\mathbb{E}[\widehat{\boldsymbol{S}}^{(\ell-2j)}]\right)^2 \le \frac{B_K}{n}.$$

因此，应用引理 24，可得存在常数 C，$C'>0$ 使得对于所有 $t\ge 0$和 $C'\le d\le n^{2/\ell}$，

$$\left\|\widehat{\boldsymbol{S}}^{(\ell-2j)}-\mathbb{E}[\widehat{\boldsymbol{S}}^{(\ell-2j)}]\right\|_{\mathrm{op}} \le \frac{Cd^{\frac{\ell}{4}-\frac{j}{2}}}{\sqrt{n}}\left(1+\frac{t^{1/2}}{n}+\left(\frac{d^{\frac{\ell}{2}-j}}{n}\right)^{1/12}t^{2/3}+\left(\frac{d^{\frac{\ell}{2}-j}}{n}\right)^{1/4}t\right),$$

以至少 $1-3d^{\frac{\ell}{2}-j}e^{-t}$ 的概率。特别地，我们可以选择 $t=d^c$，其中 $c>0$ 是一个充分小的常数，使得

$$\max_{j\in\{1,\dots,\ell/2\}}\|\boldsymbol{\Delta}_j\|_{\mathrm{op}} \le \frac{C'd^{\frac{\ell}{4}-\frac{1}{2}}}{\sqrt{n}} \lesssim \frac{1}{\sqrt{d}},$$

以至少 $1-e^{-d^c}$ 的概率。

步骤 3：的集中性。现在，我们利用适用于具有重尾元素的样本协方差矩阵的 Bai-Yin 型不等式来控制项 $\boldsymbol{\Delta}_0$。将 $\mathcal{T}(\boldsymbol{y})$ 分解为 $\mathcal{T}(\boldsymbol{y})=\mathcal{T}_+(\boldsymbol{y})-\mathcal{T}_-(\boldsymbol{y})$，其中 $\mathcal{T}_+(\boldsymbol{y})=\max\{\mathcal{T}(\boldsymbol{y})\ 0\}$，且 $\mathcal{T}_-(\boldsymbol{y})=\max\{-\mathcal{T}(\boldsymbol{y})\ 0\}$。由三角不等式以及 $|b^{(d)}_{a\ a\ ,,0}|=\Theta_d(1)$ 这一事实，只需控制这两个矩阵的集中性

$$\boldsymbol{\Delta}_{0,\pm}:=\frac{1}{n}\sum_{i=1}^{n}\mathcal{T}_{\pm}(\boldsymbol{y}_i)\boldsymbol{h}_{d,\ell/2}(\boldsymbol{z}_i)\boldsymbol{h}_{d,\ell/2}(\boldsymbol{z}_i)^{\mathsf{T}}-\mathbb{E}[\mathcal{T}_{\pm}(\boldsymbol{y})\boldsymbol{h}_{d,\ell/2}(\boldsymbol{z})\boldsymbol{h}_{d,\ell/2}(\boldsymbol{z})^{\mathsf{T}}].$$

我们重点关注矩阵$\boldsymbol{\Delta}_{0,\ +}$；$\boldsymbol{\Delta}_{0,\ -}$ 也可以用类似的方法处理。对于任意确定性单位向量$\boldsymbol{u}\in\mathbb{R}^{d^{\ell/2}}$，根据球面上的超收缩性（30）和假设4，(K1)，

$$\mathbb{E}[\langle\mathcal{T}_+^{1/2}(\boldsymbol{y}_i)\boldsymbol{h}_{d,\ell/2}(\boldsymbol{z}),\boldsymbol{u}\rangle^{2p}]\le B_K^{p/2}(2p-1)^{p\ell/2},$$

对于任意整数 $p\ge 1$。特别地，这意味着 $\langle\mathcal{T}_+^{1/2}(\boldsymbol{y})\boldsymbol{h}_{d,\ \ell/2}(\boldsymbol{z})\ \boldsymbol{u}\rangle$ 是一个次 Weibull 随机变量，因此对于任意 $0<\alpha<4/\ell\le 2$，存在一个与 n，d 无关的常数 $C>0$，使得对于任意 $t>0$，

$$\mathbb{P}\left(|\langle\mathcal{T}_+^{1/2}(\boldsymbol{y})\boldsymbol{h}_{d,\ell/2}(\boldsymbol{z}),\boldsymbol{u}\rangle|\ge t\right)\le C\exp\left(-t^{\alpha}\right).$$

根据 [GLPTJ17，定理4。7]，

$$\|\boldsymbol{\Delta}_{0,+}\|_{\mathrm{op}}\lesssim\frac{1}{n}\max_{i\in[n]}\|\mathcal{T}_+^{1/2}(\boldsymbol{y}_i)\boldsymbol{h}_{d,\ell/2}(\boldsymbol{z}_i)\|_{\mathrm{op}}^2+\sqrt{\frac{d^{\ell/2}}{n}}$$

以至少[10]的概率

$$1-8e^{-n}+\frac{1}{(10n)^4}\exp\left(-\frac{4d^{\frac{\alpha\ell}{4}}}{(3.5\ln(2d^{\ell/2}))^{2\alpha}}\right)+\frac{n^2}{2\exp\left((2nd^{\ell/2})^{\alpha/4}\right)}.$$

注意，根据假设 4.(K1) 以及 $\boldsymbol{h}_{d,\ell/2}(\boldsymbol{z})$ 的定义，我们可以确定性地得到 $\|\mathcal{T}_+^{1/2}(\boldsymbol{y}_i)\boldsymbol{h}_{d,\ell/2}(\boldsymbol{z}_i)\|_{\mathrm{op}}^2 \le \|\mathcal{H}_{d,\ell/2}(\boldsymbol{z}_i)\|_{\mathsf{F}}^2 \lesssim d^{\ell/2}$ 。因此，如果 $n \ge d^{\ell/2}$，我们已经证明存在常数 $C, c > 0$，使得对于 $C \le d \le n^{2/\ell}$,

$$\|\boldsymbol{\Delta}_{0,+}\|_{\mathrm{op}} \le C\sqrt{\frac{d^{\ell/2}}{n}},$$

以至少 $1-e^{-d^c}$ 的概率成立。对于涉及 $\mathcal{T}_-(\boldsymbol{y})$ 的第二个矩阵，也有类似的界，因此可得

$$\|\boldsymbol{\Delta}_0\|_{\mathrm{op}} \le C\sqrt{\frac{d^{\ell/2}}{n}},$$

以至少 $1-e^{-d^c}$ 的概率成立。

步骤 4：总结。 为结束证明并汇总以上得到的估计，将其分解为

$$\widehat{\boldsymbol{M}}-\mathbb{E}[\widehat{\boldsymbol{S}}]\mathbb{E}[\widehat{\boldsymbol{S}}]^{\mathsf{T}}=\left(\widehat{\boldsymbol{S}}-\mathbb{E}[\widehat{\boldsymbol{S}}]\right)\left(\widehat{\boldsymbol{S}}-\mathbb{E}[\widehat{\boldsymbol{S}}]\right)^{\mathsf{T}}+\left(\widehat{\boldsymbol{S}}-\mathbb{E}[\widehat{\boldsymbol{S}}]\right)\mathbb{E}[\widehat{\boldsymbol{S}}]^{\mathsf{T}}+\mathbb{E}[\widehat{\boldsymbol{S}}]\left(\widehat{\boldsymbol{S}}-\mathbb{E}[\widehat{\boldsymbol{S}}]\right)^{\mathsf{T}}$$

使得

$$\left\|\widehat{\boldsymbol{M}}-\mathbb{E}[\widehat{\boldsymbol{S}}]\mathbb{E}[\widehat{\boldsymbol{S}}]^{\mathsf{T}}\right\|_{\mathrm{op}} \le \|\widehat{\boldsymbol{S}}-\mathbb{E}[\widehat{\boldsymbol{S}}]\|_{\mathrm{op}}^2+2\|\mathbb{E}[\widehat{\boldsymbol{S}}]\|_{\mathrm{op}}\|\widehat{\boldsymbol{S}}-\mathbb{E}[\widehat{\boldsymbol{S}}]\|_{\mathrm{op}}.$$

根据步骤 2 和步骤 3 得到的界，以及步骤 1 中的分解，存在常数 $C, C', c > 0$，使得对所有 $C' \le d \le n^{2/\ell}$,

$$\|\widehat{\boldsymbol{S}}-\mathbb{E}[\widehat{\boldsymbol{S}}]\|_{\mathrm{op}} \le C\sqrt{\frac{d^{\ell/2}}{n}}$$

以至少为 $1-e^{-d^c}$的概率。此外，根据引理19 和假设4.(K1)， $\|\mathbb{E}[\widehat{\boldsymbol{S}}]\|_{\mathsf{F}} \le B_K^{1/2}\|\boldsymbol{\xi}_{\emptyset,\,\ell}\|_{L^2}$。结合这些估计即可完成证明。 □

B.3 多步过程：迭代张量展开

定理 4 的证明由下述引理推出，该引理控制算法2中多步过程的一个中间步骤。

引理8。 假设定理 *4* 的设置。固定 $t \in [T-1]$ 并假设在第 t 步，我们已恢复出子空间 $\widehat{\boldsymbol{U}}_{\le t} \in \mathrm{Stf}_{s\le t}(\mathbb{R}^d)$ 。令 φ 为假设 *5*中的连续模。定义

$$\delta := \varphi(\mathrm{dist}(\widehat{\boldsymbol{U}}_{\le t}, \boldsymbol{U}_{\le t})) \vee \mathrm{dist}(\widehat{\boldsymbol{U}}_{\le t}, \boldsymbol{U}_{\le t}).$$

[10]在撰写本文时， [GLPTJ17的定理4的表述中存在一个小的笔误。7], 概率中第二项的指数缺少一个负号。尽管如此，我们已与作者确认，上述表达式是正确的。

令 $\widehat{\boldsymbol{U}}_{\le t+1}$ 为算法 2在下一次迭代中返回的子空间。则存在常数C、C'、c、$c'>0$，它们仅依赖于 $\{\ell_t\}_{t\in[T]}$、s 以及假设中的常数，使得对所有 $C'\le d$、$C'\le\gamma\le\sqrt{d}$、$\delta\le 1/C'$ 和 $n\le\exp(d^{c'})$，

$$\operatorname{dist}(\widehat{\boldsymbol{U}}_{\le t+1},\boldsymbol{U}_{\le t+1})\le C\left[\frac{d^{\ell_{t+1}/2\vee 1}}{n\|\boldsymbol{\xi}_t\|_{L^2}^2}+\frac{1}{\sqrt{\gamma}}+\sqrt{\delta}\right], \tag{71}$$

概率至少为 $1-\exp(-d^c)$。

证明。为简化记号，令 $\ell:=\ell_{t+1}$。回顾一下，展开形状取为 $(a,b)=(1,0)$，若 $\ell=1$；否则取 $a=\lfloor\ell/2\rfloor$, $b=\ell-a$。令 $\boldsymbol{U}_{\le t,\perp}$ 和 $\boldsymbol{U}_{\le t,\perp}$ 分别为 $\boldsymbol{U}_{\le t}$ 和 $\boldsymbol{U}_{\le t}$的正交补。

令 $\boldsymbol{M}$ 表示在条件为 $\boldsymbol{U}_{\le t}$时于步骤 $t+1$ 构造的经验 $d_{t+1}^a\times d_{t+1}^a$ 矩阵；令 $\boldsymbol{M}$ 表示在以总体子空间 $\boldsymbol{U}_{\le t}$ 为条件时构造的对应经验矩阵（使用相应的补空间 $\boldsymbol{U}_{\le t,\perp}$ 和$\boldsymbol{U}_{\le t,\perp}$）。

步骤 1：总体矩阵。 我们首先验证 $\mathbb{E}[\boldsymbol{M}]$ 满足引理 1 的界。只需检查对称化核（限制在顶端信号方向上）K_{t+1}，其定义见（48）。固定 $\boldsymbol{V}\in\mathrm{TSym}^\ell(\mathbb{R}^{\mathsf{s}_{t+1}})$。根据定义，

$$\begin{aligned}&\mathbb{E}\left[\overline{K}_{t+1}(\boldsymbol{y}_t,\boldsymbol{y}_t')\langle\boldsymbol{V},\boldsymbol{\zeta}_{\boldsymbol{U}_{\le t},\ell}(\boldsymbol{y}_t)\rangle_{\mathsf{F}}\langle\boldsymbol{\zeta}_{\boldsymbol{U}_{\le t},\ell}(\boldsymbol{y}_t'),\boldsymbol{V}\rangle_{\mathsf{F}}\right]\\ &=\int_{\mathcal{O}_{\mathsf{s}_{\le t}}}\mathbb{E}\left[K_{t+1}(g\cdot\boldsymbol{y}_t,g\cdot\boldsymbol{y}_t')\langle\boldsymbol{V},\boldsymbol{\zeta}_{\boldsymbol{U}_{\le t},\ell}(\boldsymbol{y}_t)\rangle_{\mathsf{F}}\langle\boldsymbol{\zeta}_{\boldsymbol{U}_{\le t},\ell}(\boldsymbol{y}_t'),\boldsymbol{V}\rangle_{\mathsf{F}}\right]\mathrm{d}\pi_{\mathsf{s}_{\le t}}(g),\end{aligned}$$

其中 $g\cdot\boldsymbol{y}_t=(\boldsymbol{y},\ g\cdot\boldsymbol{U}_{\le t}^{\mathsf{T}}\boldsymbol{z})$ 和 $g\cdot\boldsymbol{y}_t'=(\boldsymbol{y}',\ g\cdot\boldsymbol{U}_{\le t}^{\mathsf{T}}\boldsymbol{z}')$。

令 $O\subseteq\mathcal{O}_{\mathsf{s}_{\le t}}$为恒等元的一个邻域，使得 $\varphi(\|\boldsymbol{U}_{\le t}-\boldsymbol{U}_{\le t}\cdot g\|_{\mathsf{F}})\le\varepsilon$对所有 $g\in O$均成立。将积分限制在 O，并对4.(K2) 在 K_{t+1}时应用假设，我们得到

$$\begin{aligned}&\mathbb{E}\left[\overline{K}_{t+1}(\boldsymbol{y}_t,\boldsymbol{y}_t')\langle\boldsymbol{V},\boldsymbol{\zeta}_{\boldsymbol{U}_{\le t},\ell}(\boldsymbol{y}_t)\rangle\langle\boldsymbol{\zeta}_{\boldsymbol{U}_{\le t},\ell}(\boldsymbol{y}_t'),\boldsymbol{V}\rangle\right]\\ &\ge\pi_{\mathsf{s}_{\le t}}(O)\mathbb{E}\left[\langle\boldsymbol{V},\boldsymbol{\zeta}_{\boldsymbol{U}_{\le t},\ell}(\boldsymbol{y}_t)\rangle^2\right]\\ &\quad-\|\boldsymbol{\xi}_{\boldsymbol{U}_{\le t},\ell}\|_{L^2}^2\int_O\mathbb{E}\left[(K_{t+1}(\boldsymbol{y}_t,\boldsymbol{y}_t')-K_{t+1}((\boldsymbol{y},g\cdot\boldsymbol{U}_{\le t}^{\mathsf{T}}\boldsymbol{z}),(\boldsymbol{y}',g\cdot\boldsymbol{U}_{\le t}^{\mathsf{T}}\boldsymbol{z}')))^2\right]^{1/2}\mathrm{d}\pi_{\mathsf{s}_{\le t}}(g).\end{aligned} \tag{72}$$

应用假设 5.(b)，我们得到

$$\mathbb{E}\left[\overline{K}_{t+1}(\boldsymbol{y}_t,\boldsymbol{y}_t')\langle\boldsymbol{V},\boldsymbol{\zeta}_{\boldsymbol{U}_{\le t},\ell}(\boldsymbol{y}_t)\rangle_{\mathsf{F}}\langle\boldsymbol{\zeta}_{\boldsymbol{U}_{\le t},\ell}(\boldsymbol{y}_t'),\boldsymbol{V}\rangle_{\mathsf{F}}\right]\ge c\mathbb{E}\left[\langle\boldsymbol{V},\boldsymbol{\zeta}_{\boldsymbol{U}_{\le t},\ell}(\boldsymbol{y}_t)\rangle_{\mathsf{F}}^2\right]-C\|\boldsymbol{\xi}_{\boldsymbol{U}_{\le t},\ell}\|_{L^2}^2\varepsilon.$$

选择 ε使第二项相较于r_t阶特征值足够小（例如，$C\|\boldsymbol{\xi}_{\boldsymbol{U}_{\le t},\ \ell}\|_{L^2}^2\varepsilon\le c\mu_{\mathsf{r}_t}/2$），引理 1 适用，从而得到常数 $c,\ C>0$ 使得对所有足够大的 d，

$$c\|\boldsymbol{\xi}_{\boldsymbol{U}_{\le t},\ell}\|_{L^2}^2\boldsymbol{\Pi}_0^{(a)}\preceq\mathbb{E}[\widehat{\boldsymbol{M}}]\preceq C\|\boldsymbol{\xi}_{\boldsymbol{U}_{\le t},\ell}\|_{L^2}^2\left(\boldsymbol{\Pi}_0^{(a)}+\frac{1}{\gamma}\boldsymbol{\Pi}_+^{(a)}+\frac{1}{\sqrt{d}}\mathbf{I}_{d^a}\right), \tag{73}$$

其中，$\boldsymbol{\Pi}_0^{(a)},\boldsymbol{\Pi}_+^{(a)}$ 按式（55）中的定义，对应于与总体 $\boldsymbol{U}_{\le t}$ 相关联的约化模型 $\nu_{d,t}$。**步骤 2：提升至 $\mathbb{R}^d$ 并讨论期望的稳定性。** 定义提升后的矩阵

$$\widehat{\boldsymbol{M}}^{(L)}:=[\widehat{\boldsymbol{U}}_{\le t,\perp}^{\otimes\ell}]\widehat{\boldsymbol{M}}[\widehat{\boldsymbol{U}}_{\le t,\perp}^{\otimes\ell}]^{\mathsf{T}}\in\mathbb{R}^{d^a\times d^a},\qquad\widetilde{\boldsymbol{M}}^{(L)}:=[\boldsymbol{U}_{\le t,\perp}^{\otimes\ell}]\widetilde{\boldsymbol{M}}[\boldsymbol{U}_{\le t,\perp}^{\otimes\ell}]^{\mathsf{T}}\in\mathbb{R}^{d^a\times d^a}.$$

$\widehat{}$根据 $\boldsymbol{M}$ 的定义，我们可以写成

$$\mathbb{E}\left[\widehat{\boldsymbol{M}}^{(L)}\right] = \mathbb{E}\left[\overline{K}_{t+1}(\hat{\boldsymbol{y}}_t, \hat{\boldsymbol{y}}_t')\mathrm{Mat}_{a,b}(\widehat{\boldsymbol{U}}_{\leq t,\perp}^{\otimes \ell}\boldsymbol{\xi}_{\widehat{\boldsymbol{U}}_{\leq t},\ell}(\hat{\boldsymbol{y}}_t))\mathrm{Mat}_{a,b}(\widehat{\boldsymbol{U}}_{\leq t,\perp}^{\otimes \ell}\boldsymbol{\xi}_{\widehat{\boldsymbol{U}}_{\leq t},\ell}(\hat{\boldsymbol{y}}_t))^{\mathsf{T}}\right],$$

其中 $\hat{\boldsymbol{y}}_t = (\boldsymbol{y}, \widehat{\boldsymbol{U}}_{\leq t}^{\mathsf{T}}\boldsymbol{z})$，且 $\hat{\boldsymbol{y}}_t'$是一个独立副本；对 $\mathbb{E}[\widetilde{\boldsymbol{M}}^{(L)}]$ 成立类似的恒等式。

我们现在对这两组期望之差给出界。利用 $\|K_{t+1}\|_\infty \leq B_K$ （假设 4.(K1)）、三角不等式和柯西–施瓦茨，

$$\begin{aligned}
&\left\|\mathbb{E}\big[\widehat{\boldsymbol{M}}^{(L)}\big] - \mathbb{E}\big[\widetilde{\boldsymbol{M}}^{(L)}\big]\right\|_{\mathsf{F}} \\
&\leq \mathbb{E}\left[\left|\overline{K}_{t+1}(\hat{\boldsymbol{y}}_t, \hat{\boldsymbol{y}}_t') - \overline{K}_{t+1}(\boldsymbol{y}_t, \boldsymbol{y}_t')\right| \|\boldsymbol{\xi}_{\widehat{\boldsymbol{U}}_{\leq t},\ell}(\hat{\boldsymbol{y}}_t)\|_{\mathsf{F}}\|\boldsymbol{\xi}_{\widehat{\boldsymbol{U}}_{\leq t},\ell}(\hat{\boldsymbol{y}}_t')\|_{\mathsf{F}}\right] \\
&\quad + B_K\mathbb{E}\left[(\|\boldsymbol{\xi}_{\widehat{\boldsymbol{U}}_{\leq t},\ell}(\hat{\boldsymbol{y}}_t)\|_{\mathsf{F}} + \|\boldsymbol{\xi}_{\boldsymbol{U}_{\leq t},\ell}(\boldsymbol{y}_t)\|_{\mathsf{F}})\|\widehat{\boldsymbol{U}}_{\leq t,\perp}^{\otimes \ell}\boldsymbol{\xi}_{\widehat{\boldsymbol{U}}_{\leq t},\ell}(\hat{\boldsymbol{y}}_t') - \boldsymbol{U}_{\leq t,\perp}^{\otimes \ell}\boldsymbol{\xi}_{\boldsymbol{U}_{\leq t},\ell}(\boldsymbol{y}_t')\|_{\mathsf{F}}\right].
\end{aligned}$$

由假设 5，若 $\delta \leq 1/2$ ，则 $\|\boldsymbol{\xi}_{\widehat{\boldsymbol{U}}_{\leq t},\ell}\|_{L^2} \leq 3\|\boldsymbol{\xi}_{\boldsymbol{U}_{\leq t},\ell}\|_{L^2}$，并利用柯西–施瓦茨，两项均由 $C\|\boldsymbol{\xi}_{\boldsymbol{U}_{\leq t},\ell}\|_{L^2}^2\delta$ 控制。因此，

$$\left\|\mathbb{E}\big[\widehat{\boldsymbol{M}}^{(L)}\big] - \mathbb{E}\big[\widetilde{\boldsymbol{M}}^{(L)}\big]\right\|_{\mathsf{F}} \leq C\|\boldsymbol{\xi}_{\boldsymbol{U}_{\leq t},\ell}\|_{L^2}^2\delta.$$

当 δ 充分小时，这意味着 $\mathbb{E}[\widehat{\boldsymbol{M}}^{(L)}]$ 的主特征子空间接近于 $\mathbb{E}[\widetilde{\boldsymbol{M}}^{(L)}]$,的主特征子空间，而由（73）它们对应于 $[\boldsymbol{U}_{\leq t,\perp}^{\otimes \ell}]\boldsymbol{\Pi}_0^{(a)}[\boldsymbol{U}_{\leq t,\perp}^{\otimes \ell}]^{\mathsf{T}}$。

步骤 3：通过 Davis–Kahan 得出结论。不等式（71）可通过重复定理 3 证明中的论证得到，其中两次应用 Davis–Kahan 正弦 Θ 定理，将经验矩阵 $\widehat{\boldsymbol{M}}$与 $\mathbb{E}\big[\widetilde{\boldsymbol{M}}^{(L)}\big]$进行比较。由此可得

$$\mathrm{dist}(\boldsymbol{U}_{\leq t,\perp}\boldsymbol{U}_{t+1}, \widehat{\boldsymbol{U}}_{\leq t,\perp}\widehat{\boldsymbol{U}}_{t+1}) \leq C\left[\frac{d^{\ell_{t+1}/2\vee 1}}{n\|\boldsymbol{\xi}_t\|_{L^2}^2} + \frac{1}{\sqrt{\gamma}} + \sqrt{\delta}\right].$$

根据三角不等式

$$\mathrm{dist}(\boldsymbol{U}_{\leq t+1}, \widehat{\boldsymbol{U}}_{\leq t+1}) \leq \mathrm{dist}(\boldsymbol{U}_{\leq t}, \widehat{\boldsymbol{U}}_{\leq t}) + \mathrm{dist}(\boldsymbol{U}_{\leq t,\perp}\boldsymbol{U}_{t+1}, \widehat{\boldsymbol{U}}_{\leq t,\perp}\widehat{\boldsymbol{U}}_{t+1})$$

证明得证。 □

注释 B.1 （ K_t的有限秩近似）。为匹配统计查询下界，并避免在构造 $\widehat{\boldsymbol{M}}$时对所有成对组合（i, j） $\in [n]^2$ 求和，我们将 K_t 替换为有限秩近似，同时保持（73）中的总体界。具体而言，我们用一个 η-网对定义 K_t 的 Haar 积分进行近似，该网位于 $\mathcal{O}_{\mathsf{s}_{\leq t}}$中，记为 $\{g_j\}_{j\in[N_\eta]}$，并定义

$$\overline{K}_t^{(\eta)} = \frac{1}{N_\eta}\sum_{j\in[N_\eta]} K_t(g_j\cdot\boldsymbol{y}_t, g_j\cdot\boldsymbol{y}_t').$$

由于 K_t满足假设 4。(K3)，核 $K_t^{(\eta)}$的秩至多为$\mathsf{m}N_\eta \leq \mathsf{m}(1+2/\eta)^{\mathsf{s}_{\leq t}^2}$，这里用到了一个标准的覆盖数界。最后，与式（72）中相同的论证表明，可以将 η选得足够小——其取值仅取决于问题中的常数——从而使关键的总体界（73）在将 K_t 替换为 $K_t^{(\eta)}$ 后仍然成立。

B.4 算法的运行时间

算法 1 中占主导地位的计算成本来自计算经验矩阵 $\widehat{\boldsymbol{M}} \in \mathbb{R}^{d^a \times d^a}$的前 t 个特征向量。我们使用子空间幂迭代来计算这些特征向量。具体而言，经过$O(\log(d^a/\varepsilon)/g)$ 次迭代，即可得到顶端 t维特征子空间的 ε近似，其中

$$g := 1 - \hat{\lambda}_{\mathsf{t}+1}/\hat{\lambda}_{\mathsf{t}},$$

$\hat{\ }$和 λ_j 表示 $\boldsymbol{M}$的第 j大特征值。

令 $\mathsf{T}_{\mathrm{mult}}$ 表示计算一次矩阵-向量乘积 $\boldsymbol{M}\boldsymbol{v}$所需的时间，其中$\boldsymbol{v} \in \mathbb{R}^{d^a}$。子空间幂迭代的每次迭代都涉及 t 次此类乘积，以及一次开销为 $O(\mathsf{t}^3)$的正交归一化步骤。因此，第一步特征分解的总运行时间为

$$O\left((\mathsf{t}\,\mathsf{T}_{\mathrm{mult}} + \mathsf{t}^3)\, \frac{\log(d/\varepsilon)}{g} \right).$$

T该算法的第二步计算矩阵的前 s_0 个特征向量

$$\sum_{s=1}^{\mathsf{t}} \mathrm{Mat}_{1,a-1}(\widehat{\boldsymbol{v}}_s)\, \mathrm{Mat}_{1,a-1}(\widehat{\boldsymbol{v}}_s)^{\mathsf{T}},$$

其中 $\{\widehat{\boldsymbol{v}}_s\}_{s=1}^{\mathsf{t}}$ 是 $\boldsymbol{M}$的主特征向量。应用相同的子空间迭代论证，此步骤需要

$$O\left(\mathsf{t}\, d^a\, \frac{\log(d/\varepsilon)}{h} \right)$$

次运算，其中

$$h := 1 - \widetilde{\mu}_{\mathsf{s}_0+1}/\widetilde{\mu}_{\mathsf{s}_0},$$

且 $\widetilde{\mu}_j$ 表示上述矩阵的第 j大特征值。在定理3的假设下，我们有 $h \geq c$以高概率成立。因此，只要$\mathsf{T}_{\mathrm{mult}} = \Omega(d^a)$，这第二步的渐近运行时间就比第一步更短，且不会影响整体运行时间。

矩阵-向量乘法。现在我们来界定计算矩阵-向量$_{\mathrm{mult}}$ 的

乘积所需的成本 $\boldsymbol{M}$。在有限秩核假设（假设4.(K3)）下，矩阵 $\boldsymbol{M}$可分解为

$$\widehat{\boldsymbol{M}} = \left(1 + \frac{\delta_{a \neq b}}{n-1}\right) \sum_{r=1}^{\mathsf{m}} \left[\boldsymbol{S}_r \boldsymbol{S}_r^{\mathsf{T}} - \delta_{a \neq b}\, \boldsymbol{D}_r \right],$$

其中

$$\boldsymbol{S}_r := \frac{1}{n} \sum_{i \in [n]} \mathcal{T}_r(\boldsymbol{y}_i)\, \mathrm{Mat}_{a,b}\big(\mathcal{H}_{d,\ell}(\boldsymbol{z}_i)\big), \qquad \boldsymbol{D}_r := \frac{1}{n^2} \sum_{i \in [n]} \mathcal{T}_r(\boldsymbol{y}_i)^2\, \mathrm{Mat}_{a,b}\big(\mathcal{H}_{d,\ell}(\boldsymbol{z}_i)\big) \mathrm{Mat}_{a,b}\big(\mathcal{H}_{d,\ell}(\boldsymbol{z}_i)\big)^{\mathsf{T}}.$$

因此，计算 $\boldsymbol{M}\boldsymbol{v}$ 可归结为计算 $O(\mathsf{m}n)$ 个涉及调和张量 Mat 的矩阵-向量乘积$_{a,\ b}(\mathcal{H}_{d,\ \ell}(\boldsymbol{z}_i))$。根据 [JKMS25，引理 15], ，每个此类乘积均可通过 $O(d^a + d^b)$ 次基本运算计算得出，无需显式构造完整矩阵。因此，

$$\mathsf{T}_{\mathrm{mult}} = O\big(\mathsf{m}n(d^a + d^b)\big).$$

C 应用的详细内容与证明

本节将第5节概述的应用具体化，并提供必要的细节和证明。特别是，我们将展示高斯和方向性 MIM 如何纳入球面对称框架，并刻画具有输入与输出之间多项式关系的球面多指标模型的学习复杂度。

C.1 高斯和方向性 MIM 的特化

本节将说明我们的框架如何应用于高斯和方向性 MIM，如第 5.1 节和第 5.2 节所述，并补充缺失的证明。我们采用的方法遵循 [JKMS25] 中提出的策略，并依赖于厄米特张量的若干性质，以及其中介绍的厄米特张量的调和分解。在进行证明之前，我们先回顾这些性质。

C.1.1 厄米特张量及其调和分解

令 $\gamma_d(\boldsymbol{x}) = (2\pi)^{-\frac{d}{2}} e^{-\|\boldsymbol{x}\|_2^2/2}$ 表示 $\mathbb{R}^d$ 中标准高斯向量的概率密度。对于任意 $d \geq 1$ 和 $k \geq 0$，定义 $\mathrm{He}_k : \mathbb{R}^d \to \mathrm{Sym}_k(\mathbb{R}^d)$ 为 $\mathbb{R}^d$ 中归一化的 k 次厄米特张量：

$$\mathrm{He}_k(\boldsymbol{x}) := \frac{(-1)^k}{\sqrt{k!}} \frac{\nabla^k \gamma_d(\boldsymbol{x})}{\gamma_d(\boldsymbol{x})}, \tag{74}$$

其中，$\nabla^k \gamma_d(\boldsymbol{x})$ 表示 γ_d 的第 k 阶导数，并将其视为一个对称的 k 张量。特别地，当 $d = 1$ 时，集合 $\{\mathrm{He}_k\}_{k \geq 0}$ 退化为 $L^2(\mathbb{R}, \gamma_1)$ 中经典的厄米特多项式正交归一基。

埃尔米特张量在 $\mathrm{Sym}_k(\mathbb{R}^d)$ 与第 k 阶维纳混沌之间实现经典等距同构：对于所有 $\boldsymbol{A} \in \mathrm{Sym}_k(\mathbb{R}^d)$ 和 $\boldsymbol{B} \in \mathrm{Sym}_j(\mathbb{R}^d)$，

$$\mathbb{E}_{\gamma_d}[\langle \boldsymbol{A}, \mathrm{He}_k(\boldsymbol{x})\rangle_{\mathsf{F}} \langle \boldsymbol{B}, \mathrm{He}_j(\boldsymbol{x})\rangle_{\mathsf{F}}] = \delta_{jk} \langle \boldsymbol{A}, \boldsymbol{B}\rangle_{\mathsf{F}}.$$

因此，任意 $f \in L^2(\gamma_d)$ 都有维纳混沌展开

$$f(\boldsymbol{x}) = \sum_{k=0}^{\infty} \langle \boldsymbol{A}_k, \mathrm{He}_k(\boldsymbol{x})\rangle_{\mathsf{F}}, \qquad \boldsymbol{A}_k := \mathbb{E}_{\gamma_d}[f(\boldsymbol{x})\mathrm{He}_k(\boldsymbol{x})] \in \mathrm{Sym}_k(\mathbb{R}^d),$$

在 $L^2(\gamma_d)$ 中收敛。与引理 20 的第二部分类似，如果此外 f 对某个 $\mathcal{O}_d^{\boldsymbol{W}}$不变，则 $\boldsymbol{W} \in \mathrm{Stf}_k(\mathbb{R}^d$可写为 $\boldsymbol{A}_k = \boldsymbol{W}^{\otimes k} \boldsymbol{B}_k$，其中某个 $\boldsymbol{B}_k \in \mathrm{Sym}_k(\mathbb{R}^s$。我们还回顾一下对所有 $\boldsymbol{w} \in \mathbb{S}^{d-1}$均成立的恒等式，

$$\mathrm{He}_k(\langle \boldsymbol{w}, \boldsymbol{x}\rangle) = \langle \boldsymbol{w}^{\otimes k}, \mathrm{He}_k(\boldsymbol{x})\rangle_{\mathsf{F}}. \tag{75}$$

更一般地，式（75）结合（标量）埃尔米特生成函数可推出

$$\exp\left(\langle \boldsymbol{x}, \boldsymbol{y}\rangle - \frac{\|\boldsymbol{y}\|_2^2}{2}\right) = \sum_{k=0}^{\infty} \frac{\langle \mathrm{He}_k(\boldsymbol{x}), \boldsymbol{y}^{\otimes k}\rangle_{\mathsf{F}}}{\sqrt{k!}},$$

因此，对于任意 $\boldsymbol{W} \in \mathrm{Stf}_k(\mathbb{R}^d)$ 和 $\boldsymbol{x} \in \mathbb{R}^d$，

$$(\boldsymbol{W}^{\mathsf{T}})^{\otimes k} \mathrm{He}_k(\boldsymbol{x}) = \mathrm{He}_k(\boldsymbol{W}^{\mathsf{T}} \boldsymbol{x}). \tag{76}$$

为了明确厄米特张量的代数结构，回顾一维埃尔米特多项式的经典闭式表达式会很方便。具体而言，对于 $u \in \mathbb{R}$ 和 $k \geq 0$,

$$\mathrm{He}_k(u) = \frac{1}{\sqrt{k!}} \sum_{j=0}^{\lfloor k/2 \rfloor} \frac{(-1)^j k!}{2^j j!(k-2j)!} u^{k-2j}.$$

结合恒等元 (75)，可得到多元埃尔米特张量的显式张量表示:

$$\mathrm{He}_k(\boldsymbol{x}) = \frac{1}{\sqrt{k!}} \sum_{j=0}^{\lfloor k/2 \rfloor} \frac{(-1)^j k!}{2^j j!(k-2j)!} \mathsf{P}_{\mathsf{sym}}\big(\boldsymbol{x}^{\otimes(k-2j)} \otimes \mathbf{I}_d^{\otimes j}\big). \tag{77}$$

下面的引理给出了埃尔米特张量到调和张量的显式分解。相关分解见 [JKMS25 的命题 2 和引理 3]。为完整起见，我们在此给出证明，该证明直接基于本文建立的张量框架，并可简短推导得出。

引理 9 （厄米特-调和分解）。 对于任意整数 $d \geq 1$ 和 $k \geq 0$，以及任意 $\boldsymbol{x} \in \mathbb{R}^d$，我们有如下分解

$$\mathrm{He}_k(\boldsymbol{x}) = \sum_{j=0}^{\lfloor k/2 \rfloor} \beta_{k,k-2j}^{(d)}(r)\, \mathsf{P}_{\mathsf{sym}} \left(\mathcal{H}_{d,k-2j}(\boldsymbol{z}) \otimes \mathbf{I}_d^{\otimes j} \right),$$

其中 $\boldsymbol{x} = r\boldsymbol{z}$ 满足 $r = \|\boldsymbol{x}\|_2$ 和 $\boldsymbol{z} = \boldsymbol{x}/\|\boldsymbol{x}\|_2$，且

$$\beta_{k,k-2j}^{(d)}(r) = \frac{\sqrt{k!}}{\kappa_{d,k-2j}\sqrt{N_{d,k-2j}}} \sum_{i=0}^{j} \frac{(-1)^i r^{k-2i}}{2^i i!(k-2i)!} f_{k-2i,j-i}^{(d)}$$

其中 $\kappa_{d,\,\ell}$，$N_{d,\,\ell}$ 定义于 (20) 以及 $f_{\ell}^{(d)}{}_{,\,j}$ 定义于引理 *16*。

此外，对于任意固定整数 ℓ、$j \geq 0$,

$$\mathbb{E}_{r\sim\chi_d}[\beta_{\ell+2j,\ell}^{(d)}(r)]^2 = \Theta_d\left(d^{-2j}\right), \quad \mathbb{E}_{r\sim\chi_d}[\beta_{\ell+2j,\ell}^{(d)}(r)^2] = \Theta_d\left(d^{-j}\right).$$

此外，存在一个常数 $C > 0$，使得对于所有 d、$j \geq C$ 以及任意固定整数 $\ell \geq 0$,

$$\frac{\mathbb{E}_{r\sim\chi_d}[\beta_{\ell+2(j+1),\ell}^{(d)}(r)]^2}{\mathbb{E}_{r\sim\chi_d}[\beta_{\ell+2j,\ell}^{(d)}(r)]^2} \leq 1, \quad \textit{and} \quad \frac{\mathbb{E}_{r\sim\chi_d}[\beta_{\ell+2(j+1),\ell}^{(d)}(r)^2]}{\mathbb{E}_{r\sim\chi_d}[\beta_{\ell+2j,\ell}^{(d)}(r)^2]} \leq 1.$$

证明。 由（77）以及对称化算子的线性性,

$$\mathrm{He}_k(\boldsymbol{x}) = \frac{1}{\sqrt{k!}} \sum_{j=0}^{\lfloor k/2 \rfloor} \frac{(-1)^j k! r^{k-2j}}{2^j j!(k-2j)!} \mathsf{P}_{\mathsf{sym}} \left(\boldsymbol{z}^{\otimes(k-2j)} \otimes \mathbf{I}_d^{\otimes j} \right),$$

其中我们写下了 $\boldsymbol{x} = r\boldsymbol{z}$，连同 $r = \|\boldsymbol{x}\|_2$ 与 $\boldsymbol{z} = \boldsymbol{x}/\|\boldsymbol{x}\|_2$。根据Fischer分解（94），我们进一步分解

$$\mathrm{He}_k(\boldsymbol{x}) = \frac{1}{\sqrt{k!}} \sum_{j=0}^{\lfloor k/2 \rfloor} \frac{(-1)^j k! r^{k-2j}}{2^j j!(k-2j)!} \sum_{i=0}^{\lfloor (k-2j)/2 \rfloor} f_{k-2j,i}^{(d)} \mathsf{P}_{\mathsf{sym}} \left(\mathsf{P}_{\mathsf{tf}}(\boldsymbol{z}^{\otimes(k-2j-2i)}) \otimes \mathbf{I}_d^{\otimes(j+i)} \right).$$

将关于 j 与 i 的求和重新指标化，令 $m=j+i$，我们得到

$$\mathrm{He}_k(\boldsymbol{x})=\sum_{m=0}^{\lfloor k/2\rfloor}\left[\frac{1}{\sqrt{k!}}\sum_{j=0}^{m}\frac{(-1)^j k! r^{k-2j}}{2^j j!(k-2j)!}f^{(d)}_{k-2j,m-j}\right]\mathsf{P}_{\mathsf{sym}}\left(\mathsf{P}_{\mathsf{tf}}(\boldsymbol{z}^{\otimes(k-2m)})\otimes\mathbf{I}_d^{\otimes m}\right).$$

根据 $\mathcal{H}_{d,\ k-2m}(\boldsymbol{z})$ 在（20）中的定义，我们可以将其改写为

$$\mathrm{He}_k(\boldsymbol{x})=\sum_{m=0}^{\lfloor k/2\rfloor}\left[\frac{\sqrt{k!}}{\kappa_{d,k-2m}\sqrt{N_{d,k-2m}}}\sum_{j=0}^{m}\frac{(-1)^j r^{k-2j}}{2^j j!(k-2j)!}f^{(d)}_{k-2j,m-j}\right]\mathsf{P}_{\mathsf{sym}}\left(\mathcal{H}_{d,k-2m}(\boldsymbol{z})\otimes\mathbf{I}_d^{\otimes m}\right).$$

至此，便完成了引理第一部分的证明。

对于第二部分，我们将一般展开式特化为 $k=\ell+2j$ 和 $m=j$。利用引理16中的 Fischer 系数显式表达式，

$$f^{(d)}_{\ell+2j-2i,j-i}=\frac{(\ell+2j-2i)!}{2^{2(j-i)}(j-i)!\ell!(d/2+\ell)_{j-i}},$$

我们得到

$$\begin{aligned}\beta^{(d)}_{\ell+2j,\ell}(r)&=\frac{\sqrt{(\ell+2j)!}}{\kappa_{d,\ell}\sqrt{N_{d,\ell}}}\sum_{i=0}^{j}\frac{(-1)^i r^{\ell+2j-2i}}{2^i i!(\ell+2j-2i)!}\frac{(\ell+2j-2i)!}{2^{2(j-i)}(j-i)!\,\ell!\,(d/2+\ell)_{j-i}}\\&=\frac{(-1)^j\sqrt{(\ell+2j)!}\,r^\ell}{\kappa_{d,\ell}\sqrt{N_{d,\ell}}\,2^j\ell!}\sum_{t=0}^{j}\frac{(-1)^t}{t!(j-t)!}\frac{(r^2/2)^t}{(d/2+\ell)_t},\end{aligned}$$

其中最后一步进行了变量替换 $t=j-i$。利用广义拉盖尔多项式的标准恒等式，

$$L_j^{(\alpha)}(x)=\frac{(\alpha+1)_j}{j!}\sum_{t=0}^{j}\binom{j}{t}\frac{(-x)^t}{(\alpha+1)_t},\qquad\alpha>-1,$$

令 $\alpha=d/2+\ell-1$，我们得到

$$\beta^{(d)}_{\ell+2j,\ell}(r)=\frac{(-1)^j\sqrt{(\ell+2j)!}}{2^j\kappa_{d,\ell}\sqrt{N_{d,\ell}}\,\ell!\,(d/2+\ell)_j}\,r^\ell\,L_j^{(d/2+\ell-1)}\left(\frac{r^2}{2}\right).$$

令 $r\sim\chi_d$ 表示一个具有 d 个自由度的卡方分布随机变量。注意到$y=r^2/2$ 服从形状参数为 $d/2$ 、尺度参数为 1 的伽马分布，即其概率密度为 $y^{d/2-1}e^{-y}/\Gamma(d/2)$，其中 $y>0$。利用积分恒等式 $\int_0^\infty y^{\beta-1}e^{-y}L_n^{(\alpha)}(y)\mathrm{d}y=\frac{\Gamma(\beta)(\alpha+1-\beta)_n}{n!}$（参见例如 [Buc69，第 119 页，公式 4β]）以及拉盖尔多项式的正交性（参见例如 [Buc69，第 136 页，公式 9]），可得

$$\mathbb{E}\left[r^\ell L_j^{(d/2+\ell-1)}\left(\frac{r^2}{2}\right)\right]=2^{\ell/2}\frac{\Gamma(d/2+\ell/2)}{\Gamma(d/2)}\frac{(\ell/2)_j}{j!},\quad\mathbb{E}\left[r^{2\ell}\left(L_j^{(d/2+\ell-1)}\left(\frac{r^2}{2}\right)\right)^2\right]=2^\ell\frac{(d/2)_{\ell+j}}{j!}.$$

将这些公式代入 $\beta^{(d)}_{\ell+2j,\ell}(r)$的表达式中，可得到显式公式

$$\mathbb{E}[\beta^{(d)}_{\ell+2j,\ell}(r)]=\frac{(-1)^j\,2^{-j+\ell/2}\sqrt{(\ell+2j)!}}{\kappa_{d,\ell}\sqrt{N_{d,\ell}}\,\ell!\,(d/2+\ell)_j}\,\frac{\Gamma(d/2+\ell/2)}{\Gamma(d/2)}\,\frac{(\ell/2)_j}{j!}$$

and

$$\mathbb{E}[\beta^{(d)}_{\ell+2j,\ell}(r)^2] = \frac{2^{\ell-2j}(\ell+2j)!}{\kappa^2_{d,\ell}\, N_{d,\ell}\,(\ell!)^2}\frac{(d/2)_\ell}{(d/2+\ell)_j}\,\frac{1}{j!}.$$

由于在 $d\to\infty$ 的情况下，$N_{d,\ell}=\Theta_d(d^\ell)$和$\kappa_{d,\ell}=\Theta_d(1)$，且 ℓ, j 固定，因此可得 t

$$\mathbb{E}[\beta^{(d)}_{\ell+2j,\ell}(r)]^2 = \Theta_d\left(d^{-2j}\right), \quad \mathbb{E}[\beta^{(d)}_{\ell+2j,\ell}(r)^2] = \Theta_d\left(d^{-j}\right).$$

为证明引理的最后一部分，我们计算以下比值

$$\frac{\mathbb{E}[\beta^{(d)}_{\ell+2(j+1),\ell}(r)]^2}{\mathbb{E}[\beta^{(d)}_{\ell+2j,\ell}(r)]^2} = \frac{(\ell+2j+2)(\ell+2j+1)(\ell/2+j)^2}{4(d/2+\ell+j)^2(j+1)^2}$$

and

$$\frac{\mathbb{E}[\beta^{(d)}_{\ell+2(j+1),\ell}(r)^2]}{\mathbb{E}[\beta^{(d)}_{\ell+2j,\ell}(r)^2]} = \frac{(\ell+2j+2)(\ell+2j+1)}{4(d/2+\ell+j)(j+1)}.$$

当 d 足够大（取决于 ℓ）时，随着 $j\to\infty$，两个比值最终都会从下方趋近于 1，由此即可完成证明。

□

Lemma 9 给出如下混沌-调和分解：对于任意 $\boldsymbol{A}\in$对称$_k(\mathbb{R}^d$),

$$\langle \boldsymbol{A}, \mathrm{He}_k(\boldsymbol{x})\rangle_{\mathsf{F}} = \sum_{j=0}^{\lfloor k/2\rfloor} \beta^{(d)}_{k,k-2j}(r)\langle \mathsf{P}_{\mathsf{tf}}(\tau^j(\boldsymbol{A})), \mathcal{H}_{d,k-2j}(\boldsymbol{z})\rangle_{\mathsf{F}}. \tag{78}$$

C.1.2 高斯 MIM

考虑一个链接函数为 $\rho\in\mathcal{P}(\mathcal{Y}\times\mathbb{R}^s)$的高斯多指标模型，即

$$(y,\boldsymbol{x})\sim \mathbb{P}^{\boldsymbol{W}_*}_\rho: \qquad \boldsymbol{x}\sim\gamma_d, \qquad y\mid \boldsymbol{x}\sim\rho(\,\cdot\mid \boldsymbol{W}_*^{\mathsf{T}}\boldsymbol{x}),$$

并令 $\boldsymbol{W}_*\in$Stf$_s(\mathbb{R}^d)$表示植入子空间。为简洁起见，以下省略对 $\boldsymbol{W}_*$ 的依赖。

该模型是球面对称不变的。因此，如示例1中所讨论的，利用极分解 $\boldsymbol{x}=r\boldsymbol{z}$ （高斯向量的），其中 $r=\|\boldsymbol{x}\|_2\sim\nu_d^R=\chi_d$ 与 $\boldsymbol{z}=\boldsymbol{x}/\|\boldsymbol{x}\|_2\sim\tau_d$ 相互独立，我们可通过定义 $\boldsymbol{y}=(y,r)$将该模型等价地视为球面多指标模型，并且

$$\boldsymbol{z}\sim\tau_d, \qquad \boldsymbol{y}=(y,r)\mid \boldsymbol{z}\sim\nu_d(\mathrm{d}\boldsymbol{y}\mid \boldsymbol{W}_*^{\mathsf{T}}\boldsymbol{z}) := \rho(\mathrm{d}y\mid r\,\boldsymbol{W}_*^{\mathsf{T}}\boldsymbol{z})\,\chi_d(\mathrm{d}r).$$

在该模型中，存在两种相互竞争的零分布概念，从而导致生成指数的两种不同定义。首先，在本文中如同 [JKMS25], 中一样，我们考虑这样一种零分布：标签与输入的径向分量同其方向解耦。具体而言，令 $\nu_d^{(Y,R)}$ 表示（y， r）在 $\mathbb{P}_{\nu_d}$下的边缘分布，并定义零分布

$$\mathbb{P}_{\nu_d,\emptyset} := \nu_d^{(Y,R)}\otimes\tau_d.$$

在此零模型下，似然比具有调和分解

$$\frac{\mathrm{d}\mathbb{P}_{\nu_d}}{\mathrm{d}\mathbb{P}_{\nu_d,\emptyset}}(y,\boldsymbol{x}) = \sum_{\ell=0}^{\infty}\langle \boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}), \mathcal{H}_{d,\ell}(\boldsymbol{z})\rangle_{\mathsf{F}}, \quad \boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}) = \mathbb{E}[\mathcal{H}_{d,\ell}(\boldsymbol{z})\mid y,r]. \tag{79}$$

或者，在近期大量关于高斯多指标模型的文献中 [DPVLB24,DLB25,JKMS25]，零分布被设定为使标签与输入完全解耦。具体而言，定义$\tilde{\mathbb{P}}_{\nu_d,\emptyset} := \nu_d^Y \otimes \chi_d \otimes \tau_d$，其中 ν_d^Y表示 y的边缘分布。在此零分布下，似然比具有维纳混沌展开

$$\frac{\mathrm{d}\mathbb{P}_{\nu_d}}{\mathrm{d}\tilde{\mathbb{P}}_{\nu_d,\emptyset}}(y,\boldsymbol{x}) = \sum_{k=0}^{\infty} \langle \boldsymbol{\psi}_{\emptyset,k}(y), \mathrm{He}_k(\boldsymbol{x}) \rangle_{\mathsf{F}}, \quad \boldsymbol{\psi}_{\emptyset,k}(y) = \mathbb{E}[\mathrm{He}_k(\boldsymbol{x}) \mid y].$$

该生成指数定义为

$$k_* := \arg\min_{k \geq 1} \Big\{ k : \|\boldsymbol{\psi}_{\emptyset,k}(y)\|^2_{L^2(\nu_d^Y)} > 0 \Big\},$$

使得 $\boldsymbol{\psi}_{\emptyset,k}(y) \equiv 0$ 对所有 $k < k_*$均成立。注意，$\|\boldsymbol{\psi}_{\emptyset,k}(y)\|_{L^2(\nu_d^Y)}$ 仅取决于链接函数 ρ，与环境维数 d 无关。事实上，由于$\boldsymbol{x} \mid y$ 的条件分布在稳定子 $\mathcal{O}_d^{\boldsymbol{W}_*}$作用下保持不变，张量 $\boldsymbol{\psi}_{\emptyset,k}(y) = \mathbb{E}[\mathrm{He}_k(\boldsymbol{x}) \mid y]$ 必定属于 $(\boldsymbol{W}_*)^{\otimes k}$的张成空间。然后，我们利用 (76) 写出

$$\boldsymbol{\psi}_{\emptyset,k}(y) = \mathbb{E}[(\boldsymbol{W}_*^{\mathsf{T}})^{\otimes k} \mathrm{He}_k(\boldsymbol{x}) \mid y] = \mathbb{E}[\mathrm{He}_k(\boldsymbol{W}_*^{\mathsf{T}} \boldsymbol{x}) \mid y]$$

其中 $\boldsymbol{W}_{\mathsf{T}_*}\, \boldsymbol{x} \sim \mathcal{N}(0,\ \mathbf{I}_{\mathsf{S}})$ 与 d 独立。

利用混沌-调和分解（78），我们可以将这个似然比改写为调和基下的形式：

$$\frac{\mathrm{d}\mathbb{P}_{\nu_d}}{\mathrm{d}\tilde{\mathbb{P}}_{\nu_d,\emptyset}}(y,\boldsymbol{x}) = \sum_{\ell=0}^{\infty} \sum_{j=0}^{\infty} \beta^{(d)}_{\ell+2j,\ell}(r) \langle \tau^j(\boldsymbol{\psi}_{\emptyset,\ell+2j}(y)), \mathcal{H}_{d,\ell}(\boldsymbol{z}) \rangle_{\mathsf{F}} = \sum_{\ell=0}^{\infty} \langle \bar{\boldsymbol{\xi}}_{\emptyset,\ell}(\boldsymbol{y}), \mathcal{H}_{d,\ell}(\boldsymbol{z}) \rangle_{\mathsf{F}}, \tag{80}$$

其中

$$\bar{\boldsymbol{\xi}}_{\emptyset,\ell}(\boldsymbol{y}) = \sum_{j \geq 0} \beta^{(d)}_{\ell+2j,\ell}(r) \mathsf{P}_{\mathsf{tf}}(\tau^j(\boldsymbol{\psi}_{\emptyset,\ell+2j}(y))). \tag{81}$$

写作

$$\frac{\mathrm{d}\mathbb{P}_{\nu_d}}{\mathrm{d}\tilde{\mathbb{P}}_{\nu_d,\emptyset}}(y,\boldsymbol{x}) = \frac{\mathrm{d}\mathbb{P}_{\nu_d}}{\mathrm{d}\mathbb{P}_{\nu_d,\emptyset}}(y,\boldsymbol{x}) \frac{\mathrm{d}\mathbb{P}_{\nu_d,\emptyset}}{\mathrm{d}\tilde{\mathbb{P}}_{\nu_d,\emptyset}}(y,\boldsymbol{x}) = \frac{\mathrm{d}\mathbb{P}_{\nu_d}}{\mathrm{d}\mathbb{P}_{\nu_d,\emptyset}}(y,\boldsymbol{x}) \frac{\mathrm{d}\nu_d^{(Y,R)}}{\mathrm{d}\nu_d^Y \otimes \nu_d^R}(y,r)$$

和 c 比较式（79）和（80）中的调和分解，我们得到如下关系

$$\bar{\boldsymbol{\xi}}_{\emptyset,\ell}(\boldsymbol{y}) = \boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}) \frac{\mathrm{d}\nu_d^{(Y,R)}}{\mathrm{d}\nu_d^Y \otimes \nu_d^R}(y,r). \tag{82}$$

由于 $\boldsymbol{\xi}_{\emptyset,0}(\boldsymbol{y}) = 1, \mathsf{P}_{\mathsf{tf}}(c) = c$ 对于任意 $c \in \mathbb{R}$，由式（81）和（82）可得

$$\frac{\mathrm{d}\nu_d^{(Y,R)}}{\mathrm{d}\nu_d^Y \otimes \nu_d^R}(y,r) = \bar{\boldsymbol{\xi}}_{\emptyset,0}(\boldsymbol{y}) = \sum_{j \geq 0} \beta^{(d)}_{2j,0}(r) \tau^j(\boldsymbol{\psi}_{\emptyset,2j}(y)) = \sum_{j \geq 0} \beta^{(d)}_{2j,0}(r) \tau^j(\boldsymbol{\phi}_{\emptyset,2j}(y)). \tag{83}$$

事实上，在较弱的假设下，系数 $\bar{\boldsymbol{\xi}}_{\emptyset,\ell}(\boldsymbol{y})$ 和 $\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y})$ 在 L^2 范数下很接近。这正是下一个引理的内容。

Lemma 10. 令 ℓ，$\mathsf{s} \geq 1$。假设存在常数 $C', L, \eta > 0$ 使得 at

$$\|\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y})\|_{L^{2+\eta}(\nu_d^{(Y,R)})} \leq L \|\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y})\|_{L^2(\nu_d^{(Y,R)})}$$

对所有 $d \geq C'$均成立。此外，假设存在常数 $C > \sqrt{\mathsf{s}} \vee \sqrt{\mathsf{s}(\lceil 2(2+\eta)/\eta \rceil - \bar{1})}$ 以及$J > 0$，使得$\|\boldsymbol{\psi}_{\emptyset,\ j}(y)\|_{L^2(\nu_d^Y)} \leq C^{-j/2}$ 对所有 $j \geq J$ 及所有 $d \geq C'$均成立。那么，存在常数$C'' > 0$，使得对所有 $d \geq C''$，

$$\frac{1}{2}\|\bar{\boldsymbol{\xi}}_{\emptyset,\ell}\|^2_{L^2(\nu_d^Y \otimes \nu_d^R)} \leq \|\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y})\|^2_{L^2(\nu_d^{(Y,R)})} \leq 2\|\bar{\boldsymbol{\xi}}_{\emptyset,\ell}\|^2_{L^2(\nu_d^Y \otimes \nu_d^R)}.$$

证明。 我们在全文中假设 d足够大，使得引理的所有假设均成立。我们还假设 $J = 0$ 为简化起见，一般情形可通过调整常数得到。由（82），

$$\|\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y})\|^2_{L^2(\nu_d^{(Y,R)})} = \mathbb{E}_{\nu_d^Y \otimes \nu_d^R}\left[\langle \boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y}), \bar{\boldsymbol{\xi}}_{\emptyset,\ell}(\boldsymbol{y})\rangle_{\mathsf{F}}\right].$$

借鉴 [JKMS25附录 G], 令

$$S = \sum_{j \geq 1} \beta^{(d)}_{2j,0}(r)\tau^j(\boldsymbol{\psi}_{\emptyset,2j}(y)) \in \mathbb{R}$$

使得 $1 + S = \mathrm{d}\nu_d^{(Y,R)} \underline{\mathrm{d}\nu_d^Y \otimes \nu_d^R}(y,\ r)$；由（83）并分解

$$\mathbb{E}_{\nu_d^Y \otimes \nu_d^R}\left[\langle \boldsymbol{\xi}_{\emptyset,\ell}, \bar{\boldsymbol{\xi}}_{\emptyset,\ell}\rangle_{\mathsf{F}}\right] = \|\bar{\boldsymbol{\xi}}_{\emptyset,\ell}\|^2_{L^2(\nu_d^Y \otimes \nu_d^R)} - \mathbb{E}_{\nu_d^Y \otimes \nu_d^R}\left[S\langle \boldsymbol{\xi}_{\emptyset,\ell}, \bar{\boldsymbol{\xi}}_{\emptyset,\ell}\rangle_{\mathsf{F}}\right].$$

由 H¨older 不等式，对于任意 $p,\ q,\ r \geq 1$ 使得 $1/p + 1/q + 1/r = 1$,

$$|\mathbb{E}_{\nu_d^Y \otimes \nu_d^R}\left[S\langle \boldsymbol{\xi}_{\emptyset,\ell}, \bar{\boldsymbol{\xi}}_{\emptyset,\ell}\rangle_{\mathsf{F}}\right]| \leq \|S\|_{L^p(\nu_d^Y \otimes \nu_d^R)}\|\bar{\boldsymbol{\xi}}^{1-1/r}_{\emptyset,\ell}\|_{L^q(\nu_d^Y \otimes \nu_d^R)}\|\boldsymbol{\xi}_{\emptyset,\ell}\|^2_{L^{2r}(\nu_d^{(Y,R)})}.$$

令 $p = q = 2(2+\eta)/\eta$和 $r = (2+\eta)/2$。那么，根据假设，$\|\boldsymbol{\xi}_{\emptyset,\ \ell}\|^2_{L^{2r}(\nu_d^{(Y,R)})} \leq L\|\boldsymbol{\xi}_{\emptyset,\ \ell}\|^2_{L^2(\nu_d^{(Y,R)})}$其中某个常数 $L > 0$ 与 d独立。此外，

$$\|\bar{\boldsymbol{\xi}}^{1-1/r}_{\emptyset,\ell}\|_{L^q(\nu_d^Y \otimes \nu_d^R)} = \|\bar{\boldsymbol{\xi}}_{\emptyset,\ell}\|^{\frac{\eta}{2+\eta}}_{L^2(\nu_d^Y \otimes \nu_d^R)} \leq \left\|\frac{\mathrm{d}\mathbb{P}_{\nu_d}}{\mathrm{d}\tilde{\mathbb{P}}_{\nu_d,\emptyset}}\right\|^{\frac{\eta}{2+\eta}}_{L^2(\tilde{\mathbb{P}}_{\nu_d,\emptyset})}$$

with

$$\left\|\frac{\mathrm{d}\mathbb{P}_{\nu_d}}{\mathrm{d}\tilde{\mathbb{P}}_{\nu_d,\emptyset}}\right\|_{L^2(\tilde{\mathbb{P}}_{\nu_d,\emptyset})} = \sum_{k=0}^{\infty}\|\boldsymbol{\psi}_{\emptyset,k}(y)\|^2_{L^2(\nu_d^Y)} \leq \sum_{k \geq 0}^{\infty} C^{-2k} < \infty$$

根据假设，与 d独立。因此，要完成证明，只需证明当 d 足够大时，$\|S\|_{L^p(\nu_d^Y \otimes \nu_d^R)}$ 可以任意小。令 $p' = \lceil 2(2+\eta)/\eta \rceil$ 为一个整数。由 H¨older 不等式,

$$\|S\|_{L^{p'}(\nu_d^Y \otimes \nu_d^R)} \leq \sum_{j \geq \lceil k_*/2 \rceil}\|\beta^{(d)}_{2j,0}(r)\|_{L^{p'}(\nu_d^R)}\|\tau^j(\boldsymbol{\psi}_{\emptyset,2j}(y))\|_{L^{p'}(\nu_d^Y)}.$$

暂时固定 j后，我们分别对右侧的两项进行界定。首先，由引理 9 可知，$\beta^{(d)}_{2j\,0,\ }(r)$ 是一个关于 $2j$ 的 r 次多项式，且仅含偶数次项。由于$r = \|\boldsymbol{x}\|_2$满足 $\boldsymbol{x} \sim \mathcal{N}(0,\ \mathbf{I}_d)$，我们可以将 $\beta^{(d)}_{2j\,0,\ }(r)$ 视为关于 $2j$ 维标准高斯向量的 $\mathbb{R}^d$ 次多项式。因此，根据高斯超收缩性（例如参见 [BLM13, 推论 5.21]），

$$\|\beta^{(d)}_{2j,0}(r)\|_{L^{p'}(\nu_d^R)} \leq (p'-1)^j\|\beta^{(d)}_{2j,0}(r)\|_{L^2(\nu_d^R)}.$$

接下来，由于 $\psi_{2j}(y)$ 的支撑集仅涉及有限个混沌层级，再次根据高斯超收缩性可得

$$\|\tau^j(\boldsymbol{\psi}_{\emptyset,2j}(y))\|_{L^{p'}(\nu_d^Y)} \le \mathsf{s}^{j/2}\|\boldsymbol{\psi}_{\emptyset,2j}(y)\|_{L^{p'}(\nu_d^Y)} \le [\sqrt{\mathsf{s}}(p'-1)]^j\|\boldsymbol{\psi}_{\emptyset,2j}(y)\|_{L^2(\nu_d^Y)}$$

综合上述两个界，我们得到

$$\|S\|_{L^{p'}(\nu_d^Y\otimes\nu_d^R)} \le \sum_{j\ge\lceil k_*/2\rceil}[\sqrt{\mathsf{s}}(p'-1)^2]^j\|\beta_{2j,0}^{(d)}(r)\|_{L^2(\nu_d^R)}\|\boldsymbol{\psi}_{\emptyset,2j}(y)\|_{L^2(\nu_d^Y)}.$$

根据假设，存在常数 $\alpha\in(0,1)$，使得 $[\sqrt{\mathsf{s}}(p'-1)^2]^j\|\boldsymbol{\psi}_{\emptyset,2j}(y)\|_{L^2(\nu_d^Y)}\le\alpha^j$ 对所有 j。因此，利用关于 $\mathbb{E}[\beta_{2j,0}^{(d)}(r)^2]$之比的估计，由引理 9，我们有

$$\frac{[\sqrt{\mathsf{s}}(p'-1)^2]^{j+1}\|\beta_{2j+2,0}^{(d)}(r)\|_{L^2(\nu_d^R)}\|\boldsymbol{\psi}_{\emptyset,2j+2}(y)\|_{L^2(\nu_d^Y)}}{[\sqrt{\mathsf{s}}(p'-1)^2]^j\|\beta_{2j,0}^{(d)}(r)\|_{L^2(\nu_d^R)}\|\boldsymbol{\psi}_{\emptyset,2j}(y)\|_{L^2(\nu_d^Y)}} \le \alpha<1.$$

因此，由此可得

$$\|S\|_{L^{p'}(\nu_d^Y\otimes\nu_d^R)} \le \frac{\alpha^{\lceil k_*/2\rceil}}{1-\alpha}\|\beta_{2\lceil k_*/2\rceil,0}^{(d)}(r)\|_{L^2(\nu_d^R)} \lesssim d^{-\lceil k_*/2\rceil/2},$$

其中最后一个不等式由引理9中 $\mathbb{E}[\beta_{2j,0}^{(d)}(r)^2]$ 的渐近行为得出。因此，对所有足够大的 d，$\|S\|_{L^{p'}(\nu_d^Y\otimes\nu_d^R)} \le 1/2$。 □

现在我们可以证明本节的主引理，它将调和系数$\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y})$的大小与生成指数 k_*联系起来。为此，我们引入收缩指标：对 $1\le\ell\le k_*$,

$$j_*(\ell)=\arg\min_{j\ge0}\left\{\left\|\mathsf{P}_{\mathsf{tf}}\left(\tau^{\lceil\frac{k_*-\ell}{2}\rceil+j}(\boldsymbol{\psi}_{\emptyset,\ell+2\lceil(k_*-\ell)/2\rceil+2j}(y))\right)\right\|_{L^2(\nu_d^Y)}^2>0\right\}. \tag{84}$$

换言之，$j_*(\ell)$ 是使收缩后的张量 ℓ 在阶数为 $\boldsymbol{\psi}_\emptyset$ 的调和张量上的投影$\ell+2\lceil(k_*-\ell)/2\rceil+2j(y)$ 非零所需的最少收缩次数。注意，根据引理 21，$j_*(k_*)=0$对于所有 d 大于某个依赖于 k_*的常数的s 都成立21。

引理 11。设 $1\le\ell\le k_*$ 且假设引理 *10* 的条件成立。那么，

$$\|\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y})\|_{L^2(\nu_d^{(Y,R)})}^2 \asymp d^{-\lceil(k_*-\ell)/2\rceil-j_*(\ell)}.$$

引理 11 的证明。再次，我们始终假设 d 足够大，使引理的所有假设都成立；为简便起见，假设引理 10 中的条件对 $J=0$ 成立。由引理 10，$\|\bar{\boldsymbol{\xi}}_{\emptyset,\ell}\|_{L^2(\nu_d^Y\otimes\nu_d^R)}^2 \asymp \|\boldsymbol{\xi}_{\emptyset,\ell}(\boldsymbol{y})\|_{L^2(\nu_d^{(Y,R)})}^2$。因此，只需证明$\|\bar{\boldsymbol{\xi}}_{\emptyset,\ell}\|_{L^2(\nu_d^Y\otimes\nu_d^R)}^2$ 的渐近行为符合所需结果。令 $m_*=\lceil(k_*-\ell)/2\rceil+j_*(\ell)$。将 $\bar{\boldsymbol{\xi}}_{\emptyset,\ell}$ 的定义代入式 (81)，并展开平方，

$$\begin{aligned}\|\bar{\boldsymbol{\xi}}_{\emptyset,\ell}\|_{L^2(\nu_d^Y\otimes\nu_d^R)}^2 &= \sum_{j,j'\ge0}\mathbb{E}_{\nu_d^Y}\left[\langle\mathsf{P}_{\mathsf{tf}}(\tau^j(\boldsymbol{\psi}_{\ell+2j}(y))),\mathsf{P}_{\mathsf{tf}}(\tau^{j'}(\boldsymbol{\psi}_{\ell+2j'}(y)))\rangle_{\mathsf{F}}\right]\mathbb{E}_{\nu_d^R}\left[\beta_{\ell+2j,\ell}^{(d)}(r)\beta_{\ell+2j',\ell}^{(d)}(r)\right]\\ &= \sum_{j\ge m_*}\mathbb{E}_{\nu_d^Y}\left[\|\mathsf{P}_{\mathsf{tf}}(\tau^j(\boldsymbol{\psi}_{\ell+2j}(y)))\|_{\mathsf{F}}^2\right]\mathbb{E}_{\nu_d^R}\left[\beta_{\ell+2j,\ell}^{(d)}(r)^2\right]\end{aligned}$$

$$+2\sum_{j'>j\geq m_*}\mathbb{E}_{\nu_d^Y}\left[\langle \mathsf{P}_{\mathsf{tf}}(\tau^j(\boldsymbol{\psi}_{\ell+2j}(y))),\mathsf{P}_{\mathsf{tf}}(\tau^{j'}(\boldsymbol{\psi}_{\ell+2j'}(y)))\rangle_{\mathsf{F}}\right]\mathbb{E}_{\nu_d^R}\left[\beta^{(d)}_{\ell+2j,\ell}(r)\beta^{(d)}_{\ell+2j',\ell}(r)\right].$$

我们先分析对角项。由于 $\boldsymbol{\psi}_{\ell+2j}(y)=\boldsymbol{W}_*^{\otimes(\ell+2j)}\boldsymbol{\phi}_{\ell+2j}(y)$，其中某个 $\boldsymbol{\phi}_{\ell+2j}(y)\in\mathsf{Sym}^{\ell+2j}(\mathbb{R}^{\mathsf{s}})$，且 P_{tf} 是正交投影，

$$\|\mathsf{P}_{\mathsf{tf}}(\tau^j(\boldsymbol{\psi}_{\ell+2j}(y)))\|_{\mathsf{F}}^2\leq\|\tau^j(\boldsymbol{\psi}_{\ell+2j}(y))\|_{\mathsf{F}}^2\leq\mathsf{s}^j\|\boldsymbol{\psi}_{\ell+2j}(y)\|_{\mathsf{F}}^2.$$

令 $\alpha\in(0,1)$，根据假设，对于所有 j，$\mathsf{s}^j\|\boldsymbol{\psi}_{\ell+2j}(y)\|^2_{L^2(\nu_d^Y)}\leq\alpha^j$。那么，

$$\sum_{j\geq m_*}\mathbb{E}_{\nu_d^Y}\left[\|\mathsf{P}_{\mathsf{tf}}(\tau^j(\boldsymbol{\psi}_{\ell+2j}(y)))\|_{\mathsf{F}}^2\right]\mathbb{E}_{\nu_d^R}\left[\beta^{(d)}_{\ell+2j,\ell}(r)^2\right]\leq\sum_{j\geq m_*}\alpha^j\mathbb{E}_{\nu_d^R}\left[\beta^{(d)}_{\ell+2j,\ell}(r)^2\right].$$

事实上，由于 $\mathbb{E}[\beta^{(d)}_{\ell+2j+2,\ell}(r)^2]/\mathbb{E}[\beta^{(d)}_{\ell+2j,\ell}(r)^2]\leq 1$ 根据引理 9，右侧的级数收敛，且

$$\sum_{j\geq m_*}\alpha^j\mathbb{E}_{\nu_d^R}\left[\beta^{(d)}_{\ell+2j,\ell}(r)^2\right]\leq\frac{\alpha^{m_*}}{1-\alpha}\mathbb{E}_{\nu_d^R}\left[\beta^{(d)}_{\ell+2m_*,\ell}(r)^2\right]\lesssim d^{-m_*}.$$

最后一个不等式由引理 $\mathbb{E}[\beta^{(d)}_{\ell+2j,\ell}(r)^2]$ 的渐近行为方程9得出。另一方面，不难看出，对角项的下界不小于求和中的任意单项。特别地，

$$\sum_{j\geq m_*}\mathbb{E}_{\nu_d^Y}\left[\|\mathsf{P}_{\mathsf{tf}}(\tau^j(\boldsymbol{\psi}_{\ell+2j}(y)))\|_{\mathsf{F}}^2\right]\mathbb{E}_{\nu_d^R}\left[\beta^{(d)}_{\ell+2j,\ell}(r)^2\right]\gtrsim d^{-m_*}\mathbb{E}_{\nu_d^Y}\left[\|\mathsf{P}_{\mathsf{tf}}(\tau^{m_*}(\boldsymbol{\psi}_{\ell+2m_*}(y)))\|_{\mathsf{F}}^2\right],$$

这里用到了引理 $\mathbb{E}[\beta^{(d)}_{\ell+2j,\ell}(r)^2]=\Theta_d(d^{-j})$ 中的事实：当 9 固定时，j。现在转向交叉项。根据柯西-施瓦茨不等式，它们可以被界定为

$$\left|\sum_{j'>j\geq m_*}\mathbb{E}_{\nu_d^Y}\left[\langle \mathsf{P}_{\mathsf{tf}}(\tau^j(\boldsymbol{\psi}_{\ell+2j}(y))),\mathsf{P}_{\mathsf{tf}}(\tau^{j'}(\boldsymbol{\psi}_{\ell+2j'}(y)))\rangle_{\mathsf{F}}\right]\mathbb{E}_{\nu_d^R}\left[\beta^{(d)}_{\ell+2j,\ell}(r)\beta^{(d)}_{\ell+2j',\ell}(r)\right]\right|\leq S_{m_*+1}S_{m_*},$$

其中我们定义了

$$S_m=\sum_{j\geq m}\sqrt{\mathbb{E}_{\nu_d^Y}\left[\|\mathsf{P}_{\mathsf{tf}}(\tau^j(\boldsymbol{\psi}_{\ell+2j}(y)))\|_{\mathsf{F}}^2\right]\mathbb{E}_{\nu_d^R}\left[\beta^{(d)}_{\ell+2j,\ell}(r)^2\right]}.$$

由上述论证可知，定义 S_m 的级数对所有 m 都收敛。事实上，沿用上述记号，对于所有充分大的 d，我们有

$$S_m\leq\sum_{j\geq m}\sqrt{\alpha^j\mathbb{E}_{\nu_d^R}\left[\beta^{(d)}_{\ell+2j,\ell}(r)^2\right]}.$$

特别地，

$$\left|\sum_{j'>j\geq m_*}\mathbb{E}_{\nu_d^Y}\left[\langle \mathsf{P}_{\mathsf{tf}}(\tau^j(\boldsymbol{\psi}_{\ell+2j}(y))),\mathsf{P}_{\mathsf{tf}}(\tau^{j'}(\boldsymbol{\psi}_{\ell+2j'}(y)))\rangle_{\mathsf{F}}\right]\mathbb{E}_{\nu_d^R}\left[\beta^{(d)}_{\ell+2j,\ell}(r)\beta^{(d)}_{\ell+2j',\ell}(r)\right]\right|\lesssim d^{-m_*-\frac{1}{2}}.$$

与对角项相比，这一项的阶更低。 □

C.1.3 方向 MIM

现在我们考虑方向型（高斯）多指标模型。这是一种相关的观测模型，其中输入的径向分量不可观测，只有标签和输入的方向可用。考虑通过链接分布 $\rho \in \mathcal{P}(\mathcal{Y} \times \mathbb{R}^{\mathsf{s}})$ 定义的高斯多指标模型，如附录 C.1.2所示，并令生成指数 k_* 按之前的方式定义。

假设我们只能观测形式为 $(y,\ \boldsymbol{z}) \in \mathcal{Y} \times \mathbb{S}^{d-1}$ 的样本，其中 $\boldsymbol{z} = \boldsymbol{x}/\|\boldsymbol{x}\|_2$ 是输入的方向。因此，给定 $\boldsymbol{z}$ 时 y 的条件分布可通过对未观测到的径向分量 $r = \|\boldsymbol{x}\|_2$ 边缘化得到。为了将此模型嵌入球面多指标模型的一般框架中，我们引入一个辅助径向变量 r，其服从确定性分布 δ_1，并且与 $(y,\ \boldsymbol{z})$ 独立。于是，我们在 $\mathcal{Y} \times \mathbb{R}_{\geq 0} \times \mathbb{S}^{d-1}$ 上定义提升分布 $\mathbb{P}_{\tilde{\nu}_d}$，如下所示：

$$(y, r, \boldsymbol{z}) \sim \mathbb{P}_{\tilde{\nu}_d}, \quad \boldsymbol{z} \sim \tau_d, \quad (y, r) \mid \boldsymbol{z} \sim \mathbb{E}_{R \sim \chi_d}[\rho(\mathrm{d}y \mid R\boldsymbol{W}_*^{\mathsf{T}}\boldsymbol{z})]\delta_1(\mathrm{d}r).$$

这种提升表示不会改变 $(y,\ \boldsymbol{z})$ 的联合分布，但使我们能够应用为球面多指标模型推导的调和分析。

在此设定下，与方向 MIM 相关联的零分布$\tilde{\ }\mathbb{P}_{\tilde{\nu}_d,\ \emptyset}$ 与附录 C.1.2 中完全解耦模型所导出的零分布一致，即

$$\mathbb{P}_{\tilde{\nu}_d,\emptyset} = \nu_d^Y \otimes \delta_1 \otimes \tau_d = \tilde{\mathbb{P}}_{\tilde{\nu}_d,\emptyset}.$$

根据（80），我们有

$$\frac{\mathrm{d}\mathbb{P}_{\tilde{\nu}_d}}{\mathrm{d}\mathbb{P}_{\tilde{\nu}_d,\emptyset}}(y, \boldsymbol{z}) = \mathbb{E}_{R \sim \chi_d}\left[\frac{\mathrm{d}\mathbb{P}_{\nu_d}}{\mathrm{d}\tilde{\mathbb{P}}_{\nu_d,\emptyset}}(y, R\boldsymbol{z})\right] = \sum_{\ell=0}^{\infty} \langle \mathbb{E}_{R \sim \chi_d}[\bar{\boldsymbol{\xi}}_{\emptyset,\ell}(y, R)], \mathcal{H}_{d,\ell}(\boldsymbol{z}) \rangle_{\mathsf{F}},$$

其中

$$\mathbb{E}_{R \sim \chi_d}[\bar{\boldsymbol{\xi}}_{\emptyset,\ell}(y, R)] = \sum_{j \geq 0} \mathbb{E}_{r \sim \chi_d}[\beta^{(d)}_{\ell+2j,\ell}(r)]\mathsf{P}_{\mathsf{tf}}(\tau^j(\boldsymbol{\psi}_{\emptyset,\ell+2j}(y))).$$

另一方面，我们也可以直接对似然比 $\mathrm{d}\mathbb{P}_{\tilde{\nu}_d}/\mathrm{d}^{\mathbb{P}_{\tilde{\nu}_d,\ \emptyset}}$ 进行调和分解，得到

$$\frac{\mathrm{d}\mathbb{P}_{\tilde{\nu}_d}}{\mathrm{d}\mathbb{P}_{\tilde{\nu}_d,\emptyset}}(y, \boldsymbol{z}) = \sum_{\ell=0}^{\infty} \langle \tilde{\boldsymbol{\xi}}_{\emptyset,\ell}(y), \mathcal{H}_{d,\ell}(\boldsymbol{z}) \rangle_{\mathsf{F}}.$$

$\tilde{\ }$这里，我们使用记号 $\boldsymbol{\xi}_{\emptyset,\ \ell}(y)$，以将这些系数与原始高斯多指标模型的系数区分开。比较这两种调和分解，可得关系

$$\tilde{\boldsymbol{\xi}}_{\emptyset,\ell}(y) = \mathbb{E}_{R \sim \chi_d}[\bar{\boldsymbol{\xi}}_{\emptyset,\ell}(y, R)] = \sum_{j \geq 0} \mathbb{E}_{R \sim \chi_d}[\beta^{(d)}_{\ell+2j,\ell}(R)]\mathsf{P}_{\mathsf{tf}}(\tau^j(\boldsymbol{\psi}_{\emptyset,\ell+2j}(y))). \tag{85}$$

我们采用与引理 11 中类似的论证，根据生成指数$\tilde{\ }$和收缩指标$\boldsymbol{\xi}_{\emptyset,\ \ell}(\boldsymbol{y})$，来刻画调和 k_* 系数 $j_*(\ell)$，其中二者定义于（84）。

引理 12. 设 $1 \leq \ell \leq k_*$。假设存在常数 C'、$J > 0$ 和 $C > \sqrt{\mathsf{s}}$ 使得$\|\boldsymbol{\psi}_{\emptyset,j}(y)\|_{L^2(\nu_d^Y)} \leq C^{-j/2}$ 对所有 $j \geq J$ 和 $d \geq C'$ 均成立。那么，

$$\|\tilde{\boldsymbol{\xi}}_{\emptyset,\ell}\|^2_{L^2(\nu_d^Y)} \asymp d^{-2\lceil (k_*-\ell)/2 \rceil - 2j_*(\ell)}.$$

证明。令 $m_* = \lceil (k_* - \ell)/2 \rceil + j_*(\ell)$，其中 $j_*(\ell)$ 按照式 (84) 中的定义给出。利用式 (‖$\tilde{\boldsymbol{\xi}}_{\emptyset,\ell}$‖$^2_{L^2(\nu_d^Y)}$中的定义展开，得到

$$\|\tilde{\boldsymbol{\xi}}_{\emptyset,\ell}\|^2_{L^2(\nu_d^Y)} = \sum_{j \geq m_*} \mathbb{E}_{\nu_d^Y}\left[\|\mathsf{P}_{\mathsf{tf}}(\tau^j(\boldsymbol{\psi}_{\ell+2j}(y)))\|_{\mathsf{F}}^2\right] \mathbb{E}_{\nu_d^R}\left[\beta^{(d)}_{\ell+2j,\ell}(R)\right]^2 + 2 \sum_{j' > j \geq m_*} \mathbb{E}_{\nu_d^Y}\left[\langle \mathsf{P}_{\mathsf{tf}}(\tau^j(\boldsymbol{\psi}_{\ell+2j}(y))), \mathsf{P}_{\mathsf{tf}}(\tau^{j'}(\boldsymbol{\psi}_{\ell+2j'}(y)))\rangle_{\mathsf{F}}\right] \mathbb{E}_{\nu_d^R}\left[\beta^{(d)}_{\ell+2j,\ell}(R)\right] \mathbb{E}_{\nu_d^R}\left[\beta^{(d)}_{\ell+2j',\ell}(R)\right].$$

证明接下来沿用引理11中的相同步骤，将 $\beta^{(d)}_{\ell+2j',\ell}(r)$ 的二阶矩替换为其一阶矩的平方，并使用引理 9 中的估计。 □

C.2 球面上的多项式学习

本节研究球面不变 MIM，其中输入 $\boldsymbol{x} \in \mathbb{R}^d$ 可分解为 $\boldsymbol{x} = r\boldsymbol{z}$，其中 $r = \sqrt{d}$ 和 $\boldsymbol{z} \sim \tau_d$，且 $\boldsymbol{W}_* = [\boldsymbol{w}_{*,1}, \ldots, \boldsymbol{w}_{*,\mathsf{s}}] \in \mathsf{Stf}_{\mathsf{s}}(\mathbb{R}^d)$ 表示植入子空间。我们观察到如下形式的响应

$$y \mid \boldsymbol{x} = \nu_d(\mathrm{d}y \mid \boldsymbol{W}_*^{\mathsf{T}}\boldsymbol{x}) := f(\boldsymbol{W}_*^{\mathsf{T}}\boldsymbol{x}) + \epsilon,$$

其中 ϵ 是独立加性噪声，而 $f : \mathbb{R}^{\mathsf{s}} \to \mathbb{R}$ 是一个次数至多为 D 的非恒定多项式。将 f 写成对称张量形式，存在张量 $\{\boldsymbol{C}_j\}_{j=1}^D$，满足 $\boldsymbol{C}_j \in \mathsf{Sym}_j(\mathbb{R}^{\mathsf{s}})$，以及一个标量 $C_0 \in \mathbb{R}$，使得

$$f(\boldsymbol{t}) = C_0 + \sum_{j=1}^{D} \langle \boldsymbol{C}_j, \boldsymbol{t}^{\otimes j} \rangle_{\mathsf{F}}.$$

The 归一化 $r = \sqrt{d}$ 确保当 $d \to \infty$ 时， $\boldsymbol{x}$ 的各坐标均为一阶量 .

在本附录中，我们将类似于高斯情形 [DLB25], 证明，学习多项式 MIM 的计算复杂度由似然比的二次谐波分量所支配。我们的论证遵循 [CM20] 的策略，并与 [DLB25，命题 4]相似。与高斯多指标模型的一个关键区别在于，似然比的埃尔米特系数与维数无关（见附录C.1），而在球面 MIM 中，调和系数以一种非平凡的方式依赖于 d。因此，在球面设定中，仅证明某个给定的调和分量非零是不够的；我们还必须控制其关于 d的大小，这需要额外的工作。

由于跃迁复杂度是根据约化模型的谐波结构定义的，因此只需在对先前已恢复的子空间进行条件化之后分析似然比（见第3.3节）。令 $\boldsymbol{U} \in \mathrm{Stf}_<(\mathbb{R}^{\mathsf{s}})$ 为一个维数为 $s_{\boldsymbol{U}} < \mathsf{s}$的已恢复子空间，并将其与其像 $\boldsymbol{U} := \boldsymbol{W}_*\boldsymbol{U} \in \mathrm{Stf}_{s_{\boldsymbol{U}}}(\mathbb{R}^d)$ 视为同一对象。设 $d_{\boldsymbol{U}} = d - s_{\boldsymbol{U}}$，并令 $\boldsymbol{U}_\perp \in \mathrm{Stf}_{d_{\boldsymbol{U}}}(\mathbb{R}^d)$ 为其正交补。如（36）中所述，我们将其分解为

$$\boldsymbol{z} = \boldsymbol{U}\boldsymbol{r}_{\boldsymbol{U}} + \sqrt{1 - \|\boldsymbol{r}_{\boldsymbol{U}}\|_2^2}\,\boldsymbol{U}_\perp \boldsymbol{z}_{\boldsymbol{U}}, \qquad \boldsymbol{r}_{\boldsymbol{U}} := \boldsymbol{U}^{\mathsf{T}}\boldsymbol{z}, \qquad \boldsymbol{z}_{\boldsymbol{U}} := \frac{\boldsymbol{U}_\perp^{\mathsf{T}}\boldsymbol{z}}{\sqrt{1 - \|\boldsymbol{r}_{\boldsymbol{U}}\|_2^2}} \in \mathbb{S}^{d_{\boldsymbol{U}}-1}.$$

对 $\boldsymbol{r}_{\boldsymbol{U}}$进行条件化后，可得到约化球面 MIM（$\boldsymbol{y}_{\boldsymbol{U}}$， $\boldsymbol{z}_{\boldsymbol{U}}$）$\sim \mathbb{P}_{\nu_d,\,\boldsymbol{U}}$，其定义见（36），并具有增广响应 $\boldsymbol{y}_{\boldsymbol{U}} =$（$y$， $\boldsymbol{r}_{\boldsymbol{U}}$）。根据第 3.3 节对跃迁复杂度的刻画，只需控制该约化模型的调和系数，即可确定完整学习过程的样本复杂度和运行时间复杂度。为使问题适定，我们假设 $\mathsf{sLeap}(\nu_d)$ $< \infty$。这意味着，对于每个真子框架 $\boldsymbol{U} \subsetneq \boldsymbol{W}_*$，约化球面

MIM（$\boldsymbol{y_U}$，$\boldsymbol{z_U}$）$\sim \mathbb{P}_{\nu_d,\ U}$ 对残差方向 $\boldsymbol{z_U}$ 保持非平凡依赖；否则，约化模型将不包含任何更多信息，也无法学习到其他方向。

推导约化模型中似然比二次谐波分量大小的变分刻画将很有用。以下引理给出了这一结果。

引理 13。在上文定义的约化球面 *MIM* $(\boldsymbol{y_U}, \boldsymbol{z_U}) \sim \mathbb{P}_{\nu_d,\ U}$ 设 $\boldsymbol{W} \in \mathrm{Stf}^{\mathsf{s}-s_U}(\mathbb{R}^{d_U})$ 是 $\boldsymbol{U}_\perp^\mathsf{T}\boldsymbol{W}_*$ 的像的一组正交归一基。那么，似然比的二次谐波分量具有如下变分刻画

$$\|\boldsymbol{\xi}_{U,2}(\boldsymbol{y})\|_{L^2}^2 = \frac{(d_U+2)d_U}{2} \sup_{G:\mathbb{R}\times\mathbb{R}^{\mathsf{s}_U}\mapsto \mathrm{Sym}_2(\mathbb{R}^{\mathsf{s}-\mathsf{s}_U})} \frac{\mathbb{E}\left[\left\langle \left(\boldsymbol{W}^\mathsf{T}\boldsymbol{z_U}\boldsymbol{z_U}^\mathsf{T}\boldsymbol{W} - \frac{1}{d_U}\mathbf{I}_{\mathsf{s}-s_U}\right), \boldsymbol{G}(\boldsymbol{y_U})\right\rangle_\mathsf{F}\right]^2}{\|G(\boldsymbol{y_U})\|_{L^2}^2}$$
$$+ \frac{(d_U+2)d_U(d_U-\mathsf{s}+\mathsf{s}_U)}{2} \sup_{g:\mathbb{R}\times\mathbb{R}^{\mathsf{s}_U}\mapsto\mathbb{R}} \frac{\mathbb{E}\left[\left(\frac{1-\|\boldsymbol{W}^\mathsf{T}\boldsymbol{z_U}\|_2^2}{d_U-(\mathsf{s}-s_U)} - \frac{1}{d_U}\right) g(\boldsymbol{y_U})\right]^2}{\|g(\boldsymbol{y_U})\|_{L^2}^2},$$

其中，上确　　确界取遍所有不几乎处处为零的平方可积函数　　.

证明。首先注意，根据调和张量的定义$\mathcal{H}_{d,\ 2}(\boldsymbol{z})$（见（20）、（88）和（91））,

$$\mathcal{H}_{d,2}(\boldsymbol{z}) = \sqrt{\frac{(d+2)d}{2}}\left(\boldsymbol{z}\boldsymbol{z}^\mathsf{T} - \frac{1}{d}\mathbf{I}_d\right)$$

因此

$$\boldsymbol{\xi}_{U,2}(\boldsymbol{y}) = \sqrt{\frac{(d_U+2)d_U}{2}}\mathbb{E}\left[\boldsymbol{z_U}\boldsymbol{z_U}^\mathsf{T} - \frac{1}{d_U}\mathbf{I}_{d_U} \mid \boldsymbol{y_U}\right].$$

注意，$\boldsymbol{y_U}$仅通过 $\boldsymbol{W}\mathrm{T}_*\,\boldsymbol{U}_\perp\boldsymbol{z_U}$依赖于 $\boldsymbol{z_U}$。因此，若我们将 $\boldsymbol{U}\mathrm{T}_\perp\boldsymbol{W}_* = \boldsymbol{W}\boldsymbol{V}\mathrm{T}$分解为某个矩阵 $\boldsymbol{V} \in \mathbb{R}^{\mathsf{s}\times(\mathsf{s}-\mathsf{s}_U)}$和某个 $\boldsymbol{W} \in \mathrm{Stf}^{\mathsf{s}-s_U}(\mathbb{R}^{d_U})$，则由 $\boldsymbol{z_U}$的旋转不变性可知，$\boldsymbol{\xi}_{U,2}(\boldsymbol{y_U})$在稳定子子群 $\mathcal{O}_{d_U}^{\boldsymbol{W}}$的作用下保持不变。特别地，根据引理 18，存在 $\boldsymbol{C}_1(\boldsymbol{y_U}) \in \mathrm{Sym}_2(\mathbb{R}^{\mathsf{s}-s_U})$和 $c_2(\boldsymbol{y_U}) \in \mathbb{R}$，使得

$$\mathbb{E}[\boldsymbol{z_U}\boldsymbol{z_U}^\mathsf{T} \mid \boldsymbol{y_U}] = \boldsymbol{W}\boldsymbol{C}_1(\boldsymbol{y_U})\boldsymbol{W}^\mathsf{T} + c_2(\boldsymbol{y_U})\left(\mathbf{I}_{d_U} - \boldsymbol{W}\boldsymbol{W}^\mathsf{T}\right).$$

对两边取迹可得

$$1 = \mathrm{Tr}(\boldsymbol{C}_1(\boldsymbol{y_U})) + c_2(\boldsymbol{y_U})(d_U - (\mathsf{s}-s_U)),$$

由此得到 $c_2(\boldsymbol{y_U}) = {}^{1-}{}_{\mathrm{Tr}(\boldsymbol{C}_1(\boldsymbol{y_U}))d_U-(\mathsf{s}-s_U)}$。此外，$\boldsymbol{C}_1(\boldsymbol{y_U}) = \mathbb{E}[\boldsymbol{W}\mathrm{T}\boldsymbol{z_U}\boldsymbol{z_U}^\mathsf{T}\boldsymbol{W} \mid \boldsymbol{y_U}]$，因此

$$\boldsymbol{\xi}_{U,2}(\boldsymbol{y}) = \sqrt{\frac{(d_U+2)d_U}{2}}\boldsymbol{W}\left(\mathbb{E}[\boldsymbol{W}^\mathsf{T}\boldsymbol{z_U}\boldsymbol{z_U}^\mathsf{T}\boldsymbol{W} \mid \boldsymbol{y_U}] - \frac{1}{d_U}\mathbf{I}_{\mathsf{s}-s_U}\right)\boldsymbol{W}^\mathsf{T}$$
$$+ \sqrt{\frac{(d_U+2)d_U}{2}}\left(\frac{1-\mathbb{E}[\|\boldsymbol{W}^\mathsf{T}\boldsymbol{z_U}\|_2^2 \mid \boldsymbol{y_U}]}{d_U-(\mathsf{s}-s_U)} - \frac{1}{d_U}\right)(\mathbf{I}_{d_U} - \boldsymbol{W}\boldsymbol{W}^\mathsf{T}).$$

利用上述两个分量的正交性，我们得到

$$\|\boldsymbol{\xi}_{U,2}(\boldsymbol{y})\|_\mathsf{F}^2 = \frac{(d_U+2)d_U}{2}\|\mathbb{E}[\boldsymbol{W}^\mathsf{T}\boldsymbol{z_U}\boldsymbol{z_U}^\mathsf{T}\boldsymbol{W} \mid \boldsymbol{y_U}] - \frac{1}{d_U}\mathbf{I}_{\mathsf{s}-s_U}\|_\mathsf{F}^2$$

$$+\frac{(d_{\boldsymbol{U}}+2)d_{\boldsymbol{U}}(d_{\boldsymbol{U}}-\mathsf{s}+\mathsf{s}_{\boldsymbol{U}})}{2}\left(\frac{1-\mathbb{E}[\|\boldsymbol{W}^{\mathsf{T}}\boldsymbol{z}_{\boldsymbol{U}}\|_2^2\mid \boldsymbol{y}_{\boldsymbol{U}}]}{d_{\boldsymbol{U}}-(\mathsf{s}-s_{\boldsymbol{U}})}-\frac{1}{d_{\boldsymbol{U}}}\right)^2.$$

根据条件期望作为 L^2-投影的标准刻画，即可得出该结果。

□

现在我们可以陈述关于学习多项式球形 MIM 的计算复杂度的主要结果。

引理 14. 在上文定义的约化球面 *MIM* $(\boldsymbol{y}_{\boldsymbol{U}},\boldsymbol{z}_{\boldsymbol{U}})\sim\mathbb{P}_{\nu_d,\boldsymbol{U}}$ 中，$\|\boldsymbol{\xi}_{\boldsymbol{U},2}(\boldsymbol{y})\|_{L^2}^2\asymp 1$。

证明。 应用三角不等式和柯西-施瓦茨不等式可得如下确定性界

$$|f(\boldsymbol{W}_*^{\mathsf{T}}\boldsymbol{x})|\le|C_0|+\sum_{j=1}^{D}\|\boldsymbol{C}_j\|_{\mathsf{F}}d^{j/2}\|\boldsymbol{W}_*^{\mathsf{T}}\boldsymbol{z}\|_2^j.$$

固定 $\eta>|C_0|$ 并定义

$$\rho_\eta:=\min_{1\le j\le D}\left(\frac{\eta-|C_0|}{D\|\boldsymbol{C}_j\|_{\mathsf{F}}}\right)^{1/j}>0.$$

那么，只要 $\|\boldsymbol{W}_{*}^{\mathsf{T}}\boldsymbol{z}\|_2\le\rho_\eta/\sqrt{d}$，就有 $|f(\boldsymbol{W}_{*}^{\mathsf{T}}\boldsymbol{x})|\le\eta$。等价地，

$$\mathcal{E}_f(\eta):=\{|f(\boldsymbol{W}_*^{\mathsf{T}}\boldsymbol{x})|>\eta\}\subseteq\left\{\|\boldsymbol{W}_*^{\mathsf{T}}\boldsymbol{z}\|_2^2>\rho_\eta^2/d\right\}.$$

令 $\boldsymbol{W}\in\mathsf{Stf}_{\mathsf{s}-s_{\boldsymbol{U}}}(\mathbb{R}^{d_{\boldsymbol{U}}})$ 为 $\boldsymbol{U}_\perp^{\mathsf{T}}\boldsymbol{W}_*$ 的像的一组正交归一基，并考虑如下分解

$$\boldsymbol{W}_*^{\mathsf{T}}\boldsymbol{z}=\boldsymbol{U}\boldsymbol{r}_{\boldsymbol{U}}+\sqrt{1-\|\boldsymbol{r}_{\boldsymbol{U}}\|_2^2}\boldsymbol{V}\boldsymbol{W}^{\mathsf{T}}\boldsymbol{z}_{\boldsymbol{U}},$$

其中 $\boldsymbol{V}\in\mathbb{R}^{\mathsf{s}\times(\mathsf{s}-\mathsf{s}_{\boldsymbol{U}})}$ 满足 $\boldsymbol{U}_\perp^{\mathsf{T}}\boldsymbol{W}_*=\boldsymbol{W}\boldsymbol{V}^{\mathsf{T}}$。于是，

$$\|\boldsymbol{W}_*^{\mathsf{T}}\boldsymbol{z}\|_2^2\le\|\boldsymbol{r}_{\boldsymbol{U}}\|_2^2+(1-\|\boldsymbol{r}_{\boldsymbol{U}}\|_2^2)\|\boldsymbol{V}\|_{\mathrm{op}}^2\|\boldsymbol{W}^{\mathsf{T}}\boldsymbol{z}_{\boldsymbol{U}}\|_2^2.$$

固定 $b\in(0,\rho_\eta^2)$ 并定义 $\mathcal{E}_{\boldsymbol{r}_{\boldsymbol{U}}}(b):=\{\|\boldsymbol{r}_{\boldsymbol{U}}\|_2^2\le b/d\}$。因此，在事件 $\mathcal{E}_f(\eta)\cap\mathcal{E}_{\boldsymbol{r}_{\boldsymbol{U}}}(b)$ 上，我们有

$$\|\boldsymbol{W}^{\mathsf{T}}\boldsymbol{z}_{\boldsymbol{U}}\|_2^2\ge\frac{\rho_\eta^2-b}{(d-b)\|\boldsymbol{V}\|_{\mathrm{op}}^2},$$

它有一个严格为正的下界，且该下界与 d 无关。

由于 $f(\boldsymbol{W}_{*}^{\mathsf{T}}\boldsymbol{x})$ 无法直接观测，定义 $\mathcal{E}_y(\eta):=\{|y|>\eta\}$ 以及 $\mathcal{E}_\epsilon(\eta):=\{|\epsilon|\le\eta\}$。于是$\mathcal{E}_f(2\eta)\cap\mathcal{E}_\epsilon(\eta)\subseteq\mathcal{E}_y(\eta)$。利用引理 13 以及测试函数

$$G(\boldsymbol{y}_{\boldsymbol{U}})=\mathbf{I}_{\mathsf{s}-s_{\boldsymbol{U}}}\mathbb{1}_{\mathcal{E}_y(\eta)\cap\mathcal{E}_{\boldsymbol{r}_{\boldsymbol{U}}}(b)},$$

我们得到

$$\|\boldsymbol{\xi}_{\boldsymbol{U},2}(\boldsymbol{y})\|_{L^2}^2\ge\frac{(d_{\boldsymbol{U}}+2)d_{\boldsymbol{U}}}{2(\mathsf{s}-\mathsf{s}_{\boldsymbol{U}})}\frac{\mathbb{E}\left[\left(\|\boldsymbol{W}^{\mathsf{T}}\boldsymbol{z}_{\boldsymbol{U}}\|_2^2-\frac{\mathsf{s}-\mathsf{s}_{\boldsymbol{U}}}{d_{\boldsymbol{U}}}\right)\mathbb{1}_{\mathcal{E}_y(\eta)\cap\mathcal{E}_{\boldsymbol{r}_{\boldsymbol{U}}}(b)}\right]^2}{\mathbb{P}(\mathcal{E}_y(\eta)\cap\mathcal{E}_{\boldsymbol{r}_{\boldsymbol{U}}}(b))}.$$

利用 ϵ 与 $\boldsymbol{x}$的独立性以及上述下界，可得

$$\|\boldsymbol{\xi}_{\boldsymbol{U},2}(\boldsymbol{y})\|_{L^2}^2\ge\frac{(d_{\boldsymbol{U}}+2)d_{\boldsymbol{U}}}{2(\mathsf{s}-\mathsf{s}_{\boldsymbol{U}})}\left(\frac{\rho_\eta^2-b}{(d-b)\|\boldsymbol{V}\|_{\mathrm{op}}^2}-\frac{\mathsf{s}-\mathsf{s}_{\boldsymbol{U}}}{d_{\boldsymbol{U}}}\right)^2\mathbb{P}(\mathcal{E}_f(2\eta)\cap\mathcal{E}_{\boldsymbol{r}_{\boldsymbol{U}}}(b))^2\mathbb{P}(\mathcal{E}_\epsilon(\eta))^2.$$

这里，我们选取足够大的 η，使得 $\mathbb{P}(\mathcal{E}_\epsilon(\eta)) > 0$，并且使得上述下界为正。

另一方面，为了给 $\mathbb{P}(\mathcal{E}_f(2\eta) \cap \mathcal{E}_{\boldsymbol{r}_U}(b))$建立下界，选取 $\boldsymbol{u} \in \mathbb{S}^{s-1}$ 使得 $\boldsymbol{V}^\top \boldsymbol{u} \neq 0$（即 $\boldsymbol{u}$ 在未恢复子空间中有一个非平凡分量），并且使 $t \mapsto f(t\boldsymbol{u})$不是常数。这样的向量存在，因为我们假设响应 y 依赖于残差方向$\boldsymbol{z}_U$。定义 $\boldsymbol{a} := \boldsymbol{W}\boldsymbol{V}^\top \boldsymbol{u} \in \mathbb{R}^{d_U}$，使得 $\boldsymbol{a} \neq 0$。根据 $\boldsymbol{z}_U \sim \tau_{d_U}$的旋转不变性，

$$\langle \boldsymbol{u}, \boldsymbol{V}\boldsymbol{W}^\mathsf{T} \boldsymbol{z}_U \rangle = \langle \boldsymbol{a}, \boldsymbol{z}_U \rangle \overset{d}{=} \|\boldsymbol{a}\|_2 z, \qquad z \sim \tilde{\tau}_{d_U,1}.$$

特别地，对每一个 $c > 0$以及所有足够大的 d （依赖于 $\|\boldsymbol{a}\|_2$ 和 c），事件$\mathcal{E}_{\boldsymbol{u}}(c) := \{\langle \boldsymbol{u}, \boldsymbol{V}\boldsymbol{W}^\top \boldsymbol{z}_U \rangle \geq c\}$ 的概率有一个独立于 d的常数下界。在$\mathcal{E}_{\boldsymbol{u}}(c) \cap \mathcal{E}_{\boldsymbol{r}_U}(b)$上，我们有

$$\langle \boldsymbol{u}, \boldsymbol{W}_*^\mathsf{T} \boldsymbol{x} \rangle \geq c(1 + o(1)) - \sqrt{b}.$$

选取足够大的 c，并利用 $|f(t\boldsymbol{u})| \to \infty$（随着 $t \to \infty$），由此可得，对所有足够大的 d，$\mathcal{E}_{\boldsymbol{u}}(c) \cap \mathcal{E}_{\boldsymbol{r}_U}(b) \subseteq \mathcal{E}_f(2\eta)$。因此，

$$\mathbb{P}(\mathcal{E}_f(2\eta) \cap \mathcal{E}_{\boldsymbol{r}_U}(b)) \geq \mathbb{P}(\mathcal{E}_{\boldsymbol{u}}(c) \cap \mathcal{E}_{\boldsymbol{r}_U}(b)) > 0$$

对所有充分大的 d，该结论一致成立。最后一个不等式特别用到了 $\boldsymbol{r}_U$ 和$\boldsymbol{z}_U$的独立性。

□

D 其他技术背景

在本附录中，我们汇集了陈述和证明主要结果所需的技术背景。为求完整，我们还给出了关于对称无迹张量及其与球谐函数之间关系的非标准结果的证明。

D.1 对称与无迹张量

对于 ℓ，$d \in \mathbb{N}$，回忆一下，我们用 $(\mathbb{R}^d)^{\otimes \ell}$ 表示阶数为 ℓ 的张量空间，其上赋予 Frobenius 内积。对称 $\mathrm{Sym}^\ell(\mathbb{R}^d) \subseteq (\mathbb{R}^d)^{\otimes \ell}$ 是对称张量构成的子空间，而 $\mathsf{P}_{\mathsf{sym}} : (\mathbb{R}^d)^{\otimes \ell} \to \mathrm{Sym}^\ell(\mathbb{R}^d)$ 是到该子空间的正交投影，由 (14) 定义。注意，$\mathsf{P}_{\mathsf{sym}}$ 关于 Frobenius 内积是自伴的。

子空间 $\mathrm{TSym}^\ell(\mathbb{R}^d) \subseteq \mathrm{Sym}^\ell(\mathbb{R}^d)$ 表示无迹对称张量空间，其偏迹（见 (15)）为零。到该子空间的正交投影显式表示为

$$\mathsf{P}_{\mathsf{tf}}(\boldsymbol{A}) = \sum_{j=0}^{\lfloor \ell/2 \rfloor} h_{\ell,j}^{(d)} \mathsf{P}_{\mathsf{sym}}(\tau^j(\boldsymbol{A}) \otimes \mathbf{I}_d^{\otimes j}), \tag{86}$$

其中 $\mathbf{I}_d$ 是 $\mathbb{R}^{d\times d}$ 中的单位矩阵，τ^j 是 τ 与自身复合的第 j 次，而系数 $h_\ell^{(d)}$，j 递归定义如下

$$h_{\ell,0}^{(d)} = 1, \qquad h_{\ell,j}^{(d)} = -\frac{(\ell-2j+2)(\ell-2j+1)}{2j(d+2\ell-2j-2)} h_{\ell,j-1}^{(d)}, \quad j \geq 1. \tag{87}$$

展开这个递归式，可得闭式表达式

$$h_{\ell,j}^{(d)} = (-1)^j \frac{(d+2\ell-2j-4)!!}{(d+2\ell-4)!!} \frac{\ell!}{2^j j!(\ell-2j)!} \tag{88}$$

对任意 $j = 0, 1, \ldots\ldots, \lfloor \ell/2 \rfloor$。

引理 15。 对于任意 ℓ，$d \in \mathbb{N}$，算子 $\mathsf{P}_{\mathsf{tf}} : \mathrm{Sym}^\ell(\mathbb{R}^d) \to \mathrm{TSym}^\ell(\mathbb{R}^d)$ 定义为 (86) 是关于弗罗贝尼乌斯内积到 $\mathrm{TSym}^\ell(\mathbb{R}^d)$ 的正交投影。

证明。 首先，P_{tf} 是自伴的：对于任意 $\boldsymbol{A}$，$\boldsymbol{B} \in \mathrm{Sym}^\ell(\mathbb{R}^d)$，

$$\begin{aligned}
\langle \mathsf{P}_{\mathsf{tf}}(\boldsymbol{A}), \boldsymbol{B} \rangle_{\mathsf{F}} &= \sum_{j=0}^{\lfloor \ell/2 \rfloor} h_{\ell,j}^{(d)} \langle \tau^j(\boldsymbol{A}) \otimes \mathbf{I}_d^{\otimes j}, \boldsymbol{B} \rangle_{\mathsf{F}} \\
&= \sum_{j=0}^{\lfloor \ell/2 \rfloor} h_{\ell,j}^{(d)} \langle \tau^j(\boldsymbol{A}), \tau^j(\boldsymbol{B}) \rangle_{\mathsf{F}} = \sum_{j=0}^{\lfloor \ell/2 \rfloor} h_{\ell,j}^{(d)} \langle \boldsymbol{A}, \tau^j(\boldsymbol{B}) \otimes \mathbf{I}_d^{\otimes j} \rangle_{\mathsf{F}} = \langle \boldsymbol{A}, \mathsf{P}_{\mathsf{tf}}(\boldsymbol{B}) \rangle_{\mathsf{F}}.
\end{aligned}$$

接下来证明，对于所有 $\boldsymbol{A} \in \mathrm{Sym}^\ell(\mathbb{R}^d)$，$\mathsf{P}_{\mathsf{tf}}(\boldsymbol{A}) \in \mathrm{TSym}^\ell(\mathbb{R}^d)$。对于 $\ell = 0$，1，我们有 $\mathsf{P}_{\mathsf{tf}}(\boldsymbol{A}) = \boldsymbol{A}$，结论显然成立。假设 $\ell \geq 2$，并定义齐次多项式

$$p_{\mathsf{P}_{\mathsf{tf}}(\boldsymbol{A})}(\boldsymbol{x}) = \langle \mathsf{P}_{\mathsf{tf}}(\boldsymbol{A}), \boldsymbol{x}^{\otimes \ell} \rangle_{\mathsf{F}}.$$

由于 $\boldsymbol{x}^{\otimes \ell}$ 是对称的，

$$p_{\mathsf{P}_{\mathsf{tf}}(\boldsymbol{A})}(\boldsymbol{x}) = \sum_{j=0}^{\lfloor \ell/2 \rfloor} h_{\ell,j}^{(d)} \langle \tau^j(\boldsymbol{A}) \otimes \mathbf{I}_d^{\otimes j}, \boldsymbol{x}^{\otimes \ell} \rangle_{\mathsf{F}} = \sum_{j=0}^{\lfloor \ell/2 \rfloor} h_{\ell,j}^{(d)} \langle \tau^j(\boldsymbol{A}), \boldsymbol{x}^{\otimes (\ell-2j)} \rangle_{\mathsf{F}} \|\boldsymbol{x}\|_2^{2j}.$$

对这个多项式取拉普拉斯算子，

$$\Delta p_{\mathsf{P}_{\mathsf{tf}}(\boldsymbol{A})}(\boldsymbol{x}) = \sum_{j=0}^{\lfloor \ell/2 \rfloor} h_{\ell,j}^{(d)} \Delta \left(\langle \tau^j(\boldsymbol{A}), \boldsymbol{x}^{\otimes(\ell-2j)} \rangle_{\mathsf{F}} \|\boldsymbol{x}\|_2^{2j} \right) \tag{89}$$

with

$$\begin{aligned}
&\Delta \left(p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}) \|\boldsymbol{x}\|_2^{2j} \right) \\
&= \sum_{i=1}^{d} \partial_{x_i}^2 \left(p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}) \|\boldsymbol{x}\|_2^{2j} \right) \\
&= \Delta p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}) \|\boldsymbol{x}\|_2^{2j} + 2 \sum_{i=1}^{d} \partial_{x_i} p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}) \partial_{x_i} \|\boldsymbol{x}\|_2^{2j} + p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}) \Delta \|\boldsymbol{x}\|_2^{2j} \\
&= \Delta p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}) \|\boldsymbol{x}\|_2^{2j} + 4j \langle \nabla p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}), \boldsymbol{x} \rangle \|\boldsymbol{x}\|_2^{2(j-1)} + 2j(d + 2(j-1)) p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}) \|\boldsymbol{x}\|_2^{2(j-1)},
\end{aligned}$$

其中我们记 $p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}) = \langle \tau^j(\boldsymbol{A}), \boldsymbol{x}^{\otimes(\ell-2j)} \rangle_{\mathsf{F}}$。由于 $p_{\tau^j(\boldsymbol{A})}$ 是次数为 $\ell - 2j$ 的齐次多项式，根据欧拉齐次函数定理，

$$\langle \nabla p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}), \boldsymbol{x} \rangle = (\ell - 2j) p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}).$$

将其代入（89），得到

$$\begin{aligned}
&\Delta p_{\mathsf{P}_{\mathsf{tf}}(\boldsymbol{A})}(\boldsymbol{x}) \\
&= \sum_{j=0}^{\lfloor \ell/2 \rfloor} h_{\ell,j}^{(d)} \left(\Delta p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}) \|\boldsymbol{x}\|_2^{2j} + 4j(\ell - 2j) p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}) \|\boldsymbol{x}\|_2^{2(j-1)} + 2j(d + 2(j-1)) p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}) \|\boldsymbol{x}\|_2^{2(j-1)} \right) \\
&= \sum_{j=0}^{\lfloor \ell/2 \rfloor} h_{\ell,j}^{(d)} \Delta p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}) \|\boldsymbol{x}\|_2^{2j} + \sum_{j=1}^{\lfloor \ell/2 \rfloor} 2j h_{\ell,j}^{(d)} (2(\ell - 2j) + d + 2(j-1)) p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}) \|\boldsymbol{x}\|_2^{2(j-1)}.
\end{aligned}$$

注意到以下关系，我们可以将 $\Delta p_{\tau^j(\boldsymbol{A})}$ 与 $p_{\tau^{j+1}(\boldsymbol{A})}$ 联系起来：

$$\begin{aligned}
\frac{\partial^2}{\partial x_i^2} p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}) &= \frac{\partial^2}{\partial x_i^2} \langle \tau^j(\boldsymbol{A}), \boldsymbol{x}^{\otimes(\ell-2j)} \rangle_{\mathsf{F}} \\
&= \frac{\partial}{\partial x_i} \langle \tau^j(\boldsymbol{A}), (\ell - 2j) \boldsymbol{x}^{\otimes(\ell-2j-1)} \otimes \boldsymbol{e}_i \rangle_{\mathsf{F}} \\
&= (\ell - 2j)(\ell - 2j - 1) \langle \tau^j(\boldsymbol{A}), \boldsymbol{x}^{\otimes(\ell-2j-2)} \otimes \boldsymbol{e}_i \otimes \boldsymbol{e}_i \rangle_{\mathsf{F}}.
\end{aligned}$$

对 $i \in [d]$ 求和，并注意到 $\sum_{i=1}^{d} \boldsymbol{e}_i \otimes \boldsymbol{e}_i = \mathbf{I}_d$，得到

$$\Delta p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}) = (\ell - 2j)(\ell - 2j - 1) \langle \tau^j(\boldsymbol{A}), \boldsymbol{x}^{\otimes(\ell-2j-2)} \otimes \mathbf{I}_d \rangle_{\mathsf{F}} = (\ell - 2j)(\ell - 2j - 1) p_{\tau^{j+1}(\boldsymbol{A})}(\boldsymbol{x}). \tag{90}$$

将其代入前一个表达式，我们得到

$$\begin{aligned}
&\Delta p_{\mathsf{P}_{\mathsf{tf}}(\boldsymbol{A})}(\boldsymbol{x}) \\
&= \sum_{j=0}^{\lfloor \ell/2 \rfloor} h_{\ell,j}^{(d)} (\ell - 2j)(\ell - 2j - 1) p_{\tau^{j+1}(\boldsymbol{A})}(\boldsymbol{x}) \|\boldsymbol{x}\|_2^{2j} + 2j h_{\ell,j}^{(d)} (2(\ell - 2j) + d + 2(j-1)) p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}) \|\boldsymbol{x}\|_2^{2(j-1)}
\end{aligned}$$

$$= \sum_{j=1}^{\lfloor \ell/2 \rfloor} \left(h_{\ell,j-1}^{(d)} (\ell - 2(j-1))(\ell - 2(j-1) - 1) + 2j h_{\ell,j}^{(d)} (2(\ell - 2j) + d + 2(j-1)) \right) p_{\tau^j(\boldsymbol{A})}(\boldsymbol{x}) \|\boldsymbol{x}\|_2^{2(j-1)}.$$

由（$h_\ell^{(d)}$，$_j$ 定义的系数87）满足递推关系

$$h_{\ell,j-1}^{(d)} (\ell - 2(j-1))(\ell - 2(j-1) - 1) + 2j h_{\ell,j}^{(d)} (2(\ell - 2j) + d + 2(j-1)) = 0$$

对任意 $j \in \{1, \ldots, \lfloor \ell/2 \rfloor\}$。因此，$\Delta p_{\mathsf{P}_{\mathsf{tf}}(\boldsymbol{A})}(\boldsymbol{x}) = 0$ 对所有 $\boldsymbol{x} \in \mathbb{R}^d$。由 (90，这尤其意味着

$$0 = \Delta p_{\mathsf{P}_{\mathsf{tf}}(\boldsymbol{A})}(\boldsymbol{x}) = \ell(\ell-1) \langle \tau(\mathsf{P}_{\mathsf{tf}}(\boldsymbol{A})), \boldsymbol{x}^{\otimes(\ell-2)} \rangle_{\mathsf{F}}$$

对所有 $\boldsymbol{x} \in \mathbb{R}^d$。由于秩一张量张成对称张量空间，因此$\tau(\mathsf{P}_{\mathsf{tf}}(\boldsymbol{A})) = 0$，进而 $\mathsf{P}_{\mathsf{tf}}(\boldsymbol{A}) \in \mathrm{TSym}\ell(\mathbb{R}^d)$。

此外，如果 $\boldsymbol{A} \in \mathrm{TSym}\ell(\mathbb{R}^d)$，那么 $\tau^j(\boldsymbol{A}) = 0$ 对所有 $j \geq 1$ 都成立，进而 $\mathsf{P}_{\mathsf{tf}}(\boldsymbol{A}) = h_\ell^{(d)}$，$_0 \boldsymbol{A} = \boldsymbol{A}$。由此完成了证明：$\mathsf{P}_{\mathsf{tf}}$ 是到 $\mathrm{TSym}\ell(\mathbb{R}^d)$ 上的正交投影。 □

将 P_{tf} 的定义扩展到非对称张量上会很有用。在这种情况下，对于任意 $\boldsymbol{A} \in (\mathbb{R}^d)^{\otimes \ell}$，我们将记 $\mathsf{P}_{\mathsf{tf}}(\boldsymbol{A}) = \mathsf{P}_{\mathsf{tf}}(\mathsf{P}_{\mathsf{sym}}(\boldsymbol{A}))$。作为两个嵌套正交投影的复合，这仍然是到 $\mathrm{TSym}\ell(\mathbb{R}^d)$ 上的正交投影。

对称张量可以分解为无迹对称张量的直和，这对应于将 $\mathrm{Sym}\ell(\mathbb{R}^d)$——视为一个 $\mathcal{O}_d$ 表示——进行半单分解，得到不可约表示：

$$\mathrm{Sym}_\ell(\mathbb{R}^d) \cong \bigoplus_{j=0}^{\lfloor \ell/2 \rfloor} \mathrm{TSym}_{\ell-2j}(\mathbb{R}^d).$$

引理 16（Fischer分解）。对于任意 ℓ、$d \in \mathbb{N}$ 和 $\boldsymbol{A} \in \mathrm{Sym}\ell(\mathbb{R}^d)$，我们有如下分解

$$\boldsymbol{A} = \sum_{j=0}^{\lfloor \ell/2 \rfloor} f_{\ell,j}^{(d)} \mathsf{P}_{\mathsf{sym}} \left(\mathsf{P}_{\mathsf{tf}}(\tau^j(\boldsymbol{A})) \otimes \mathbf{I}_d^{\otimes j} \right),$$

其中

$$f_{\ell,j}^{(d)} = \frac{\ell!}{2^{2j} j! (\ell - 2j)! (d/2 + \ell - 2j)_j}.$$

证明。通过对 ℓ 进行归纳，可以直接证明存在常数 $\{f_{\ell,j}^{(d)}\}_{j=0}^{\lfloor \ell/2 \rfloor}$，使得

$$\boldsymbol{A} = \sum_{j=0}^{\lfloor \ell/2 \rfloor} f_{\ell,j}^{(d)} \mathsf{P}_{\mathsf{sym}} \left(\mathsf{P}_{\mathsf{tf}}(\tau^j(\boldsymbol{A})) \otimes \mathbf{I}_d^{\otimes j} \right)$$

对所有 $\boldsymbol{A} \in$对称群$\ell(\mathbb{R}^d)$ 成立。将其代入（86），可得

$$\begin{aligned} \mathsf{P}_{\mathsf{tf}}(\boldsymbol{A}) = \sum_{j=0}^{\lfloor \ell/2 \rfloor} h_{\ell,j}^{(d)} \mathsf{P}_{\mathsf{sym}} \left(\tau^j(\boldsymbol{A}) \otimes \mathbf{I}_d^{\otimes j} \right) &= \sum_{j=0}^{\lfloor \ell/2 \rfloor} h_{\ell,j}^{(d)} \sum_{i=0}^{\lfloor \ell/2 - j \rfloor} f_{\ell-2j,i}^{(d)} \mathsf{P}_{\mathsf{sym}} \left(\mathsf{P}_{\mathsf{tf}}(\tau^{j+i}(\boldsymbol{A})) \otimes \mathbf{I}_d^{\otimes (j+i)} \right) \\ &= \sum_{k=0}^{\lfloor \ell/2 \rfloor} \left(\sum_{j=0}^{k} h_{\ell,j}^{(d)} f_{\ell-2j,k-j}^{(d)} \right) \mathsf{P}_{\mathsf{sym}} \left(\mathsf{P}_{\mathsf{tf}}(\tau^k(\boldsymbol{A})) \otimes \mathbf{I}_d^{\otimes k} \right). \end{aligned}$$

因此，若将系数与平凡方程 $\mathsf{P}_{\mathsf{tf}}(\boldsymbol{A}) = h_{\ell\,0}^{(d)}$,$\mathsf{P}_{\mathsf{sym}}(\mathsf{P}_{\mathsf{tf}}(\boldsymbol{A}))$ 对应起来，我们得到

$$\sum_{j=0}^{k} h_{\ell,j}^{(d)} f_{\ell-2j,k-j}^{(d)} = \begin{cases} 1 & \text{if } k=0 \\ 0 & \text{if } k \geq 1. \end{cases}$$

这给出了递推关系

$$f_{\ell,0}^{(d)} = 1, \qquad f_{\ell,k}^{(d)} = -\sum_{j=1}^{k} h_{\ell,j}^{(d)} f_{\ell-2j,k-j}^{(d)}$$

关于系数 $\{f_{\ell,j}^{(d)}\}_{j=0}^{\lfloor \ell/2 \rfloor}$ 的 $\{h_{\ell,j}^{(d)}\}_{j=0}^{\lfloor \ell/2 \rfloor}$，由此我们得到引理中所述的闭式表达。

□

经典 Fischer 分解以 [Fis18], 命名，它将齐次多项式分解为调和多项式和径向分量。鉴于调和多项式与无迹对称张量之间的同构（见引理19），引理 16正是经典 Fischer 分解的张量类比。

投影 P_{tf} 作用于秩一张量时可简化为 $\boldsymbol{x}^{\otimes \ell}$，其中 $\boldsymbol{x} \in \mathbb{R}^d$ 是任意向量。根据 P_{tf} 在（86）中的定义，并利用 $\tau^j(\boldsymbol{x}^{\otimes \ell}) = \|\boldsymbol{x}\|_2^{2j} \boldsymbol{x}^{\otimes(\ell-2j)}$,

$$\mathsf{P}_{\mathsf{tf}}(\boldsymbol{x}^{\otimes \ell}) = \sum_{j=0}^{\lfloor \ell/2 \rfloor} h_{\ell,j}^{(d)} \|\boldsymbol{x}\|_2^{2j} \mathsf{P}_{\mathsf{sym}}(\boldsymbol{x}^{\otimes(\ell-2j)} \otimes \mathbf{I}_d^{\otimes j}). \tag{91}$$

由于 P_{tf}是正交投影，我们有 $\langle \mathsf{P}_{\mathsf{tf}}(\boldsymbol{x}^{\otimes \ell}), \boldsymbol{x}^{\otimes \ell} \rangle_{\mathsf{F}} = \|\mathsf{P}_{\mathsf{tf}}(\boldsymbol{x}^{\otimes \ell})\|_{\mathsf{F}}^2$。利用（91）和$\langle \mathbf{I}_d^{\otimes j}, \boldsymbol{x}^{\otimes 2j} \rangle_{\mathsf{F}} = \langle \mathbf{I}_d, \boldsymbol{x}^{\otimes 2} \rangle^j{}_{\mathsf{F}} = \|\boldsymbol{x}\|_2^{2j}$，我们可以将该量表示为

$$\|\mathsf{P}_{\mathsf{tf}}(\boldsymbol{x}^{\otimes \ell})\|_{\mathsf{F}}^2 = \sum_{j=0}^{\lfloor \ell/2 \rfloor} h_{\ell,j}^{(d)} \|\boldsymbol{x}\|_2^{2j} \langle \boldsymbol{x}^{\otimes(\ell-2j)} \otimes \mathbf{I}_d^{\otimes j}, \boldsymbol{x}^{\otimes \ell} \rangle_{\mathsf{F}} = \sum_{j=0}^{\lfloor \ell/2 \rfloor} h_{\ell,j}^{(d)} \|\boldsymbol{x}\|_2^{2\ell}. \tag{92}$$

特别地，如果 $\|\boldsymbol{x}\|_2 = 1$，那么

$$\|\mathsf{P}_{\mathsf{tf}}(\boldsymbol{x}^{\otimes \ell})\|_{\mathsf{F}}^2 = \sum_{j=0}^{\lfloor \ell/2 \rfloor} h_{\ell,j}^{(d)} = 1/\kappa_{d,\ell}^2 \tag{93}$$

最后，将引理 16 中的 Fischer分解应用于秩一张量，可得

$$\boldsymbol{x}^{\otimes \ell} = \sum_{j=0}^{\lfloor \ell/2 \rfloor} f_{\ell,j}^{(d)} \|\boldsymbol{x}\|_2^{2j} \mathsf{P}_{\mathsf{sym}} \left(\mathsf{P}_{\mathsf{tf}}(\boldsymbol{x}^{\otimes(\ell-2j)}) \otimes \mathbf{I}_d^{\otimes j} \right), \tag{94}$$

其中，系数 $f_{\ell}^{(d)}{}_{,j}$ 的定义与引理 16中相同。

D.2 张量上的正交群作用

投影算子 $\mathsf{P}_{\mathsf{sym}}$ 和 P_{tf} 在（14）和（86）中定义，它们与在张量上作用的正交群τ 在（15）中定义的偏迹算子 $\mathcal{O}_d$ 相互作用良好。

引理 17。偏迹 $\tau : \mathrm{Sym}_\ell(\mathbb{R}^d) \to \mathrm{Sym}_{\ell-2}(\mathbb{R}^d)$ 以及投影算子 $\mathsf{P}_{\mathsf{sym}} : (\mathbb{R}^d)^{\otimes \ell} \to \mathrm{Sym}_\ell(\mathbb{R}^d)$ 和 $\mathsf{P}_{\mathsf{tf}} : \mathrm{Sym}_\ell(\mathbb{R}^d) \to \mathrm{TSym}_\ell(\mathbb{R}^d)$ 相对于 $\mathcal{O}_d$ 的作用是等变的。

证明。张量上的 $\mathcal{O}_d$ 作用与指标置换及缩并可交换。因此，$\mathsf{P}_{\mathsf{sym}}$作为对所有置换取平均的平均算子，满足$\mathsf{P}_{\mathsf{sym}}(g \cdot \boldsymbol{A}) = g \cdot \mathsf{P}_{\mathsf{sym}}(\boldsymbol{A})$；而 τ作为与不变张量$\mathbf{I}_d$进行的缩并，对所有 $g \in \mathcal{O}_d$满足 $\tau(g \cdot \boldsymbol{A}) = g \cdot \tau(\boldsymbol{A})$。最后，$\mathsf{P}_{\mathsf{tf}}$是形如 $\boldsymbol{A} \mapsto \mathsf{P}_{\mathsf{sym}}(\tau^j(\boldsymbol{A}) \otimes \mathbf{I}_d^{\otimes j})$的算子的有限线性组合，其中每个算子都是等变的，因为$\mathbf{I}_d$是 $\mathcal{O}_d$-不变的，且 τ、$\mathsf{P}_{\mathsf{sym}}$都是等变的。因此，对所有 $g \in \mathcal{O}_d$，$\mathsf{P}_{\mathsf{tf}}(g \cdot \boldsymbol{A}) = g \cdot \mathsf{P}_{\mathsf{tf}}(\boldsymbol{A})$，证明完毕。 □

对于 $\boldsymbol{W} \in \mathrm{Stf}_{\mathsf{s}}(\mathbb{R}^d)$，定义 $\mathrm{Sym}_\ell^{\boldsymbol{W}}(\mathbb{R}^d) \subseteq \mathrm{Sym}_\ell(\mathbb{R}^d)$ 为由对称张量构成、且在稳定子子群作用下保持不变的子空间 $\mathcal{O}_d^{\boldsymbol{W}}$，该稳定子子群定义于 (17)，即

$$\mathrm{Sym}_\ell^{\boldsymbol{W}}(\mathbb{R}^d) := \{\boldsymbol{A} \in \mathrm{Sym}_\ell(\mathbb{R}^d) : g \cdot \boldsymbol{A} = \boldsymbol{A} \text{ for all } g \in \mathcal{O}_d^{\boldsymbol{W}}\}.$$

无迹对称张量的相应子空间记为 $\mathrm{TSym}_\ell^{\boldsymbol{W}}(\mathbb{R}^d) \subseteq \mathrm{TSym}_\ell(\mathbb{R}^d)$。对于 $\mathsf{s} = 0$，$\mathrm{Sym}_\ell^{\boldsymbol{W}}(\mathbb{R}^d)$ 是旋转不变对称张量的子空间。根据 [GW09，定理 5.3，3] 或下文的引理 18，它要么是平凡子空间 $\{0\}$ (当 ℓ 为奇数时)，要么是由 $\mathsf{P}_{\mathsf{sym}}(\mathbf{I}_d^{\otimes p})$ 张成的子空间（当 $\ell = 2p$ 为偶数时）。

子空间 $\mathrm{Sym}_\ell^{\boldsymbol{W}}(\mathbb{R}^d)$ 有一个便于使用的分解，可根据张量与由标架 $\boldsymbol{W}$ 及其正交补张成的子空间之间的相互作用，对张量的各个分量进行区分。下面的引理将对此作出精确说明。

引理 18。设 $\mathsf{s} \geq 0$, $d \geq \mathsf{s}$ 和 $\ell \geq 0$ 为整数。设 $\boldsymbol{W} \in \mathrm{Stf}_{\mathsf{s}}(\mathbb{R}^d)$ 为一个正交s标架，并设 $\boldsymbol{A} \in \mathrm{Sym}_\ell^{\boldsymbol{W}}(\mathbb{R}^d)$ 为在稳定子子群 $\mathcal{O}_d^{\boldsymbol{W}}$ 的作用下保持不变的任意对称张量。那么，存在张量 $\boldsymbol{B}_{\ell-2j} \in \mathrm{Sym}_{\ell-2j}(\mathbb{R}^d)$、$j = 0$、……、$\lfloor \ell/2 \rfloor$，使得

$$\boldsymbol{A} = \sum_{j=0}^{\lfloor \ell/2 \rfloor} \mathsf{P}_{\mathsf{sym}}(((\boldsymbol{W}\boldsymbol{W}^{\mathsf{T}})^{\otimes(\ell-2j)} \boldsymbol{B}_{\ell-2j}) \otimes (\mathbf{I}_d - \boldsymbol{W}\boldsymbol{W}^{\mathsf{T}})^{\otimes j}).$$

证明。不失一般性，设 $\boldsymbol{W} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{\mathsf{s}}]$，其中 $\{\boldsymbol{e}_i\}_{i=1}^d$ 是 $\mathbb{R}^d$的典范基。将 $\mathbb{R}^d = U \oplus V$分解为 $U =$张成空间$(\boldsymbol{e}_1$、……、$\boldsymbol{e}_{\mathsf{s}})$和 $V =$张成空间$(\boldsymbol{e}_{\mathsf{s}+1}$、……、$\boldsymbol{e}_d)$。

对于每个 $k \in \{0, \ldots, \ell\}$，令 $\boldsymbol{A}^{(k)}$ 为 $\boldsymbol{A}$ 中恰好有 $\ell - k$ 个指标属于 U、且有 k 个指标属于 V 的分量，即

$$\boldsymbol{A}^{(k)}_{i_1,\ldots,i_\ell} = \begin{cases} \boldsymbol{A}_{i_1,\ldots,i_\ell} & \text{if exactly } \ell - k \text{ of the indices } i_1, \ldots, i_\ell \text{ are in } [\mathsf{s}], \\ 0 & \text{otherwise,} \end{cases}$$

并记为 $\boldsymbol{A} = \sum_{k=0}^{\ell} \boldsymbol{A}^{(k)}$。由于 $\boldsymbol{A} \in \mathrm{Sym}_\ell^{\boldsymbol{W}}(\mathbb{R}^d)$，张量 $\{\boldsymbol{A}^{(k)}\}_{k=0}^{\ell}$ 的支撑集位于互不相交的指标集合上，且 $\mathcal{O}_d^{\boldsymbol{W}}$的作用保持属于 U和 V的指标数量不变，因此对每个 $k \in \{0, \ldots, \ell\}$，都有 $\boldsymbol{A}^{(k)} \in \mathrm{Sym}_\ell^{\boldsymbol{W}}(\mathbb{R}^d)$。

固定 $k \in \{0, \ldots, \ell\}$，并考虑分量 $\boldsymbol{A}^{(k)}$。固定 $\{\boldsymbol{u}_i\}_{i \in [\ell-k]} \subseteq U$，并考虑多重线性形式

$$F_{\boldsymbol{u}_{1:\ell-k}} : \boldsymbol{v}_{1:k} \in V^k \mapsto \langle \boldsymbol{A}^{(k)}, \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_{\ell-k} \otimes \boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_k \rangle_{\mathsf{F}},$$

其中 $\boldsymbol{u}_{1:\ell-k} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{\ell-k})$。根据构造，$F_{\boldsymbol{u}_{1:\ell-k}}$关于 $\boldsymbol{v}_{1:k}$是对称的。此外，对于任意$g \in O(V)$，我们可以将 g扩展为 $\mathcal{O}_d^{\boldsymbol{W}}$中的一个元素，让它在 U上起平凡作用。由于 $\boldsymbol{A}^{(k)}$ 在 $\mathcal{O}_d^{\boldsymbol{W}}$的作用下保持不变，因此 $F_{\boldsymbol{u}_{1:\ell-k}}$ 是 $O(V)$-不变的。根据 [GW09，定理 5.3，3], 我们得到 $F_{\boldsymbol{u}_{1:\ell-k}}$ 的如下刻画，它取决于 k的奇偶性：

(i) 如果 k 为奇数，那么对于任意选择的 $\boldsymbol{u}_{1:\ell-k}$，都有 $F_{\boldsymbol{u}_{1:\ell-k}} \equiv 0$。这意味着 $\boldsymbol{A}^{(k)} \equiv 0$。

(ii) 若 k 为偶数，则存在一个标量 $C \equiv C(\boldsymbol{u}_{1:\ell-k})$，使得

$$F_{\boldsymbol{u}_{1:\ell-k}} = C\mathsf{P}_{\mathsf{sym}}\left((\mathbf{I}_d - \boldsymbol{W}\boldsymbol{W}^\mathsf{T})^{\otimes(k/2)}\right).$$

设 k 为偶数。映射

$$C : \boldsymbol{u}_{1:\ell-k} \in U^{\ell-k} \mapsto C(\boldsymbol{u}_{1:\ell-k})$$

是定义在 U 上的对称多重线性形式。因此，存在唯一的张量 $\boldsymbol{C}_{\ell-k} \in \mathsf{Sym}^{\ell-k}(U)$，使得

$$C(\boldsymbol{u}_{1:\ell-k}) = \langle \boldsymbol{C}_{\ell-k}, \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_{\ell-k}\rangle_\mathsf{F}.$$

我们将 $\boldsymbol{C}_{\ell-k}$ 扩展为 $\mathsf{Sym}^{\ell-k}(\mathbb{R}^d)$ 中一个支撑集位于 $[\mathsf{s}]$ 上指标处的张量，定义为

$$\boldsymbol{C}_{\ell-k} = (\boldsymbol{W}\boldsymbol{W}^\mathsf{T})^{\otimes(\ell-k)}\boldsymbol{C}_{\ell-k}.$$

现在定义

$$\tilde{\boldsymbol{A}}^{(k)} := \mathsf{P}_{\mathsf{sym}}\left(\left((\boldsymbol{W}\boldsymbol{W}^\mathsf{T})^{\otimes(\ell-k)}\boldsymbol{C}_{\ell-k}\right) \otimes (\mathbf{I}_d - \boldsymbol{W}\boldsymbol{W}^\mathsf{T})^{\otimes(k/2)}\right).$$

对于任意 $\{\boldsymbol{u}_i\}_{i\in[\ell-k]} \subseteq U$ 和 $\{\boldsymbol{v}_i\}_{i\in[k]} \subseteq V$，有

$$\begin{aligned}\langle \tilde{\boldsymbol{A}}^{(k)}, \boldsymbol{u}_{1:\ell-k} \otimes \boldsymbol{v}_{1:k}\rangle_\mathsf{F} &= c_{\ell,k}\langle \boldsymbol{C}_{\ell-k}, \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_{\ell-k}\rangle_\mathsf{F}\langle \mathsf{P}_{\mathsf{sym}}\left((\mathbf{I}_d - \boldsymbol{W}\boldsymbol{W}^\mathsf{T})^{\otimes(k/2)}\right), \boldsymbol{v}_{1:k}\rangle_\mathsf{F} \\ &= c_{\ell,k}F_{\boldsymbol{u}_{1:\ell-k}}(\boldsymbol{v}_{1:k})\end{aligned}$$

对于某个常数 $c_{\ell,\,k}$ 它仅取决于 ℓ 和 k。常数 $c_{\ell,\,k}$ 是一个组合因子，它将指标 $[\ell]$ 的排列与固定指标子集 U 和 V的排列联系起来。由于 $\boldsymbol{A}^{(k)}$ 和 $\tilde{\boldsymbol{A}}^{(k)}$ 都是对称张量，前一个等式意味着 $c_{\ell,\,k}\boldsymbol{A}^{(k)} = \tilde{\boldsymbol{A}}^{(k)}$。证明至此完成。

□

D.3 无迹对称张量与球谐函数之间的同构

使用球谐函数关于无迹对称张量的规范表示会更方便。下面的引理表明，式（21）中的求值映射定义了一个 $\mathcal{O}_d$-等变等距同构，它将配备 Frobenius 内积的 $\mathsf{TSym}^\ell(\mathbb{R}^d)$ 映射到配备 $L^2(\tau_d)$ 内积的 $\mathscr{S}_{d,\,\ell}$。

引理 19。对于任意 d、$\ell \in \mathbb{N}$，由式 (21) 定义的映射 $\Phi_{d,\,\ell}$ 是 *TSym*$\mathrm{TSym}^\ell(\mathbb{R}^d)$ 与 $\mathscr{S}_{d,\,\ell}$之间的等距同构。此外，$\Phi_{d,\,\ell}$ 在 $\mathcal{O}_d$的作用下是等变的，即对于任意 $g \in \mathcal{O}_d$和 $\boldsymbol{A} \in \mathrm{TSym}^\ell(\mathbb{R}^d)$，$\Phi_{d,\,\ell}(g \cdot \boldsymbol{A}) = g \cdot \Phi_{d,\,\ell}(\boldsymbol{A})$。

证明。任取 $\boldsymbol{A} \in$TSym$^\ell(\mathbb{R}^d)$。由于 P_{tf} 是自伴的，我们可以写成

$$\Phi(\boldsymbol{A})(\boldsymbol{x}) = \kappa_{d,\ell}\sqrt{N_{d,\ell}}\langle \boldsymbol{A}, \boldsymbol{x}^{\otimes\ell}\rangle_\mathsf{F}$$

对于所有 $\boldsymbol{x} \in \mathbb{R}^d$。我们首先证明 $\Phi_{d,\,\ell}$ 是一个同构。为证明 $\Phi_{d,\,\ell}$ 定义良好，请注意，对于任意 $\boldsymbol{x} \in \mathbb{R}^d$,

$$\Delta\Phi_{d,\ell}(\boldsymbol{A})(\boldsymbol{x}) = \kappa_{d,\ell}\sqrt{N_{d,\ell}}\Delta\langle \boldsymbol{A}, \boldsymbol{x}^{\otimes\ell}\rangle_\mathsf{F} = \kappa_{d,\ell}\sqrt{N_{d,\ell}}\ell(\ell-1)\langle \tau(\boldsymbol{A}), \boldsymbol{x}^{\otimes(\ell-2)}\rangle_\mathsf{F} = 0$$

根据（90），利用 $\boldsymbol{A} \in \mathrm{TSym}_\ell(\mathbb{R}^d)$。因此，$\Phi_{d,\ell}(\boldsymbol{A})$ 是一个调和多项式，因此当限制在球面上时，$\Phi_{d,\ell}(\boldsymbol{A}) \in \mathscr{S}_{d,\ell}$。

$\Phi_{d,\ell}$ 的单射性直接源于 $\mathrm{Sym}_\ell(\mathbb{R}^d)$ 由形如 $\boldsymbol{x}^{\otimes \ell}$ 的张量张成。接下来，我们证明 $\Phi_{d,\ell}$ 是满射。为此，取 p 为 ℓ 次齐次调和多项式，且属于 $\mathbb{R}^d$，并逐项定义 $\boldsymbol{A} \in \mathrm{Sym}_\ell(\mathbb{R}^d)$:

$$\boldsymbol{A}_{i_1,\ldots,i_\ell} = \frac{1}{\ell!} \frac{\partial^\ell p}{\partial x_{i_1} \cdots \partial x_{i_\ell}}(\boldsymbol{0}).$$

注意，由于多项式的偏导数可交换，$\boldsymbol{A}$是对称的。此外，

$$p(\boldsymbol{x}) = \frac{1}{\ell!} \sum_{i_1,\ldots,i_\ell=1}^{d} \frac{\partial^\ell p}{\partial x_{i_1} \cdots \partial x_{i_\ell}}(\boldsymbol{0}) x_{i_1} \cdots x_{i_\ell} = \langle \boldsymbol{A}, \boldsymbol{x}^{\otimes \ell} \rangle_{\mathsf{F}}$$

因为 p 是 ℓ次齐次的。最后，利用 p是调和多项式这一事实，由式（90）可得

$$0 = \Delta p(\boldsymbol{x}) = \ell(\ell-1)\langle \tau(\boldsymbol{A}), \boldsymbol{x}^{\otimes(\ell-2)} \rangle_{\mathsf{F}}$$

对所有 $\boldsymbol{x} \in \mathbb{R}^d$均成立，因此 $\boldsymbol{A} \in \mathrm{TSym}_\ell(\mathbb{R}^d)$。根据构造，$\Phi_{d,\ell}(\boldsymbol{A}) = p$，因此 $\Phi_{d,\ell}$ 是满射。

最后，我们证明 $\Phi_{d,\ell}$ 是等距映射。为此，任取 $\boldsymbol{A}$，$\boldsymbol{B} \in \mathrm{TSym}_\ell(\mathbb{R}^d)$。那么，

$$\langle \Phi_{d,\ell}(\boldsymbol{A}), \Phi_{d,\ell}(\boldsymbol{B}) \rangle_{L^2(\tau_d)} = \kappa_{d,\ell}^2 N_{d,\ell} \left\langle \mathbb{E}_{\boldsymbol{x}\sim\tau_d}[\boldsymbol{x}^{\otimes 2\ell}], \boldsymbol{A} \otimes \boldsymbol{B} \right\rangle_{\mathsf{F}}.$$

根据 τ_d的旋转不变性，对每个 $\boldsymbol{Q} \in \mathcal{O}_d$,

$$\boldsymbol{Q} \cdot \mathbb{E}_{\boldsymbol{x}\sim\tau_d}[\boldsymbol{x}^{\otimes 2\ell}] = \mathbb{E}_{\boldsymbol{x}\sim\tau_d}[(\boldsymbol{Q}\boldsymbol{x})^{\otimes 2\ell}] = \mathbb{E}_{\boldsymbol{x}\sim\tau_d}[\boldsymbol{x}^{\otimes 2\ell}],$$

其中我们用了 $\boldsymbol{Q}^{\otimes \ell}\boldsymbol{x}^{\otimes 2\ell} = (\boldsymbol{Q}\boldsymbol{x})^{\otimes \ell}$。因此，$\mathbb{E}_{\boldsymbol{x}\sim\tau_d}[\boldsymbol{x}^{\otimes 2\ell}] \in \mathrm{Sym}_{2\ell}(\mathbb{R}^d)$ 在 $\mathcal{O}_d$ 的作用下保持不变。

由引理 18，存在一个常数 c 使得 $\mathbb{E}_{\boldsymbol{x}\sim\tau_d}[\boldsymbol{x}^{\otimes 2\ell}] = c\mathsf{P}_{\mathsf{sym}}(\mathbf{I}_d^{\otimes \ell})$，因此

$$\langle \Phi_{d,\ell}(\boldsymbol{A}), \Phi_{d,\ell}(\boldsymbol{B}) \rangle_{L^2(\tau_d)} = c\kappa_{d,\ell}^2 N_{d,\ell} \left\langle \mathsf{P}_{\mathsf{sym}}(\mathbf{I}_d^{\otimes \ell}), \boldsymbol{A} \otimes \boldsymbol{B} \right\rangle_{\mathsf{F}}.$$

展开内积，

$$\left\langle \mathsf{P}_{\mathsf{sym}}(\mathbf{I}_d^{\otimes \ell}), \boldsymbol{A} \otimes \boldsymbol{B} \right\rangle_{\mathsf{F}} = \frac{1}{\ell!} \sum_{\sigma \in \mathfrak{S}_{2\ell}} \sum_{i_1,\ldots,i_{2\ell}} \prod_{j=1}^{\ell} \delta_{i_{\sigma(j)}, i_{\sigma(\ell+j)}} \boldsymbol{A}_{i_1,\ldots,i_\ell} \boldsymbol{B}_{i_{\ell+1},\ldots,i_{2\ell}}.$$

如果 $\sigma \in \mathfrak{S}_{2\ell}$ 满足存在 $j \in [\ell]$ 使得 $\{\sigma(j)\ \sigma(\ell+j)\} \subseteq [\ell]$ 或 $\{\sigma(j)\ \sigma(\ell+j)\} \subseteq [2\ell]\backslash[\ell]$，那么内积 $\langle (\mathbf{I}_d^{\otimes \ell})^\sigma, \boldsymbol{A} \otimes \boldsymbol{B} \rangle_{\mathsf{F}}$ 将收缩 $\boldsymbol{A}$ 或 $\boldsymbol{B}$ 的两个指标。由于二者均为 $\boldsymbol{A}$，$\boldsymbol{B} \in \mathrm{TSym}_\ell(\mathbb{R}^d)$，这些项不会对内积有贡献。因此，对上述结果有贡献的排列 $\sigma \in \mathfrak{S}_{2\ell}$ 只有遵循将 $[2\ell]$ 划分为 $[\ell]$ 和 $[2\ell] \setminus [\ell]$的自然划分的那些。特别是，由于 $\boldsymbol{A}$，$\boldsymbol{B}$ 是对称的，这意味着

$$\left\langle \mathsf{P}_{\mathsf{sym}}(\mathbf{I}_d^{\otimes \ell}), \boldsymbol{A} \otimes \boldsymbol{B} \right\rangle_{\mathsf{F}} = c\langle \boldsymbol{A}, \boldsymbol{B} \rangle_{\mathsf{F}}$$

对于一个（可能不同的）常数 $c > 0$。综合以上各项，存在常数 $c > 0$，使得对所有 $\boldsymbol{A}, \boldsymbol{B} \in \mathrm{TSym}_\ell(\mathbb{R}^d)$,

$$\langle \Phi_{d,\ell}(\boldsymbol{A}), \Phi_{d,\ell}(\boldsymbol{B}) \rangle_{L^2(\tau_d)} = c\kappa_{d,\ell}^2 N_{d,\ell} \langle \boldsymbol{A}, \boldsymbol{B} \rangle_{\mathsf{F}}.$$

这意味着 $\mathbb{E}_{\boldsymbol{x}\sim\tau_d}[\mathsf{P}_{\mathsf{tf}}(\boldsymbol{x}^{\otimes\ell}) \otimes \mathsf{P}_{\mathsf{tf}}(\boldsymbol{x}^{\otimes\ell})] \in \mathbb{L}_{d,\ell}$ 具有谱分解

$$\mathbb{E}_{\boldsymbol{x}\sim\tau_d}[\mathsf{P}_{\mathsf{tf}}(\boldsymbol{x}^{\otimes\ell}) \otimes \mathsf{P}_{\mathsf{tf}}(\boldsymbol{x}^{\otimes\ell})] = c \sum_{j\in[N_{d,\ell}]} \boldsymbol{V}_j \otimes \boldsymbol{V}_j$$

与 $\{\boldsymbol{V}_j\}_{j\in[N_{d,\ell}]}$ TSym 的一个正交归一基$_\ell(\mathbb{R}^d)$。对等式两边取迹，并利用（92），

$$cN_{d,\ell} = \mathbb{E}_{\boldsymbol{x}\sim\tau_d}[\|\mathsf{P}_{\mathsf{tf}}(\boldsymbol{x}^{\otimes\ell})\|_{\mathsf{F}}^2] = \sum_{j=0}^{\lfloor \ell/2 \rfloor} h_{\ell,j}^{(d)} = \frac{1}{\kappa_{d,\ell}^2}$$

其中系数 $h_\ell^{(d)}$，$_j$ 定义见（88）。因此，

$$\langle \Phi_{d,\ell}(\boldsymbol{A}), \Phi_{d,\ell}(\boldsymbol{B}) \rangle_{L^2(\tau_d)} = \langle \boldsymbol{A}, \boldsymbol{B} \rangle_{\mathsf{F}}.$$

这表明 $\Phi_{d,\ \ell}$ 是保持内积的线性双射。等变性由 $\Phi_{d、\ \ell}$ 的定义以及 P_{tf} 的等变性可得（见引理17）。□

对于任意 $\mathsf{s} \in \mathbb{N}$ 和正交归一 s-标架 $\boldsymbol{W} \in \mathbb{R}^{d\times \mathsf{s}}$，令 $\mathscr{S}_d^{W}{}_{,\ \ell} \subseteq \mathscr{S}_{d,\ \ell}$ 表示在 $\mathcal{O}_d^{\boldsymbol{W}}$的作用下保持不变的球谐函数子空间。由于 $\mathcal{O}_d^{\boldsymbol{W}}$是 $\mathcal{O}_d$的稳定子子群，根据引理19 可知，$\Phi_{d,\ \ell}$ 限制为 $\mathrm{TSym}_\ell^{\boldsymbol{W}}(\mathbb{R}^d)$ 与$\mathscr{S}_{d\,\ell}^{\boldsymbol{W}}{}_{,}$ 之间的等距同构。我们得到 $L^2(\mathbb{S}^{d-1},\ \tau_d)$ 关于调和张量的一个分解。

引理 20。令 $f \in L^2$（$\mathbb{S}^{d-1}$、τ_d）为任意对象。则我们可以写成

$$f(\boldsymbol{x}) = \sum_{\ell\geq 0} \langle \boldsymbol{A}_{d,\ell}, \mathcal{H}_{d,\ell}(\boldsymbol{x}^{\otimes\ell}) \rangle_{\mathsf{F}}$$

对所有 $\boldsymbol{x} \in \mathbb{S}^{d-1}$，其中张量系数 $\boldsymbol{A}_{d,\ \ell} \in \mathrm{TSym}\ell$（$\mathbb{R}^d$）由下式唯一确定：

$$\boldsymbol{A}_{d,\ell} = \mathbb{E}_{\boldsymbol{x}\sim\tau_d}[f(\boldsymbol{x})\mathcal{H}_{d,\ell}(\boldsymbol{x}^{\otimes\ell})]$$

对所有 $\ell \geq 0$。

此外，如果 f 在某个正交归一 s-标架 $\boldsymbol{W} \in \mathbb{R}^{d\times \mathsf{s}}$的稳定子子群 $\mathcal{O}_d^{\boldsymbol{W}}$下保持不变，则对于所有 $\ell \geq 0$，有 $\boldsymbol{A}_{d、\ \ell} \in \mathit{TSym}_\ell^{\boldsymbol{W}}(\mathbb{R}^d)$，并且存在张量 $\boldsymbol{B}_{\mathsf{s}、\ \ell} \in \mathit{Sym}\ell(\mathbb{R}^{\mathsf{s}})$，使得对于所有 $\ell \geq 0$，有 $\boldsymbol{A}_{d、\ \ell} = \mathit{P}_{\mathsf{tf}}(\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s}、\ \ell})$，其中 $\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s}、\ \ell}$表示通过 $\boldsymbol{W}$将 $\boldsymbol{B}_{\mathsf{s}、\ \ell}$典范嵌入到 $\mathit{Sym}\ell(\mathbb{R}^d)$ 中。

证明。令 $\{\psi_j\}_{j=1}^{n_\ell}$ 为 $\mathscr{S}_{d、\ \ell}$ 的一个正交归一基，并将展开式写成 $L^2(\mathbb{S}^{d-1}、\ \tau_d)$

$$f(\boldsymbol{x}) = \sum_{\ell\geq 0} \sum_{j=1}^{n_\ell} \langle f, \psi_j \rangle_{L^2(\tau_d)} \psi_j(\boldsymbol{x})$$

对所有 $\boldsymbol{x} \in \mathbb{S}^{d-1}$。根据引理 19的等距同构，$L^2(\tau_d)$-正交投影$f_\ell$ 将 f 投影到 $\mathscr{S}_{d、\ \ell}$，可将其识别为唯一的张量 $\boldsymbol{A}_{d、\ \ell} \in \mathrm{TSym}\ell(\mathbb{R}^d)$，使得

$$f_\ell(\boldsymbol{x}) = \langle \boldsymbol{A}_{d,\ell}, \mathcal{H}_{d,\ell}(\boldsymbol{x}^{\otimes\ell}) \rangle_{\mathsf{F}}.$$

此外，根据 $\Phi_{d,\ell}$ 的等距映射性质，我们有

$$\langle \boldsymbol{A}_{d,\ell}, \boldsymbol{B} \rangle_{\mathsf{F}} = \langle f_\ell, \Phi_{d,\ell}(\mathsf{P}_{\mathsf{tf}}(\boldsymbol{B})) \rangle_{L^2(\tau_d)} = \mathbb{E}[f(\boldsymbol{x})\Phi_{d,\ell}(\boldsymbol{B})(\boldsymbol{x})] = \left\langle \mathbb{E}[f(\boldsymbol{x})\mathcal{H}_{d,\ell}(\boldsymbol{x}^{\otimes\ell})], \boldsymbol{B} \right\rangle_{\mathsf{F}}$$

对每个 $\boldsymbol{B} \in \mathrm{Sym}_\ell(\mathbb{R}^d)$ 都成立，其中期望是相对于 $\boldsymbol{x} \sim \tau_d$ 计算的。这里，我们用到了这样一个事实：由于 P_{tf} 是一个正交投影，对于所有 $\boldsymbol{B} \in \mathrm{Sym}_\ell(\mathbb{R}^d)$，$\Phi(\boldsymbol{B}) = \Phi(\mathsf{P}_{\mathsf{tf}}(\boldsymbol{B}))$。由此可得 $\boldsymbol{A}_{d,\ell} = \mathbb{E}[f(\boldsymbol{x})\mathcal{H}_{d,\ell}(\boldsymbol{x}^{\otimes \ell})] \in \mathrm{TSym}_\ell(\mathbb{R}^d)$。对所有 $\ell \geq 0$ 求和，即得到 f 的上述 L^2 展开式。

现在假设对于某个正交归一 s-标架 $\boldsymbol{W} \in \mathbb{R}^{d\times \mathsf{s}}$，$f$ 是 $\mathcal{O}_d^{\boldsymbol{W}}$-不变的。利用 $\mathcal{H}_d$ 的 ℓ 等变性，我们得到

$$\boldsymbol{Q} \cdot \boldsymbol{A}_{d,\ell} = \mathbb{E}[f(\boldsymbol{x})\mathcal{H}_{d,\ell}((\boldsymbol{Q}\boldsymbol{x})^{\otimes \ell})] = \mathbb{E}[f(\boldsymbol{Q}^{\mathsf{T}}\boldsymbol{x})\mathcal{H}_{d,\ell}(\boldsymbol{x}^{\otimes \ell})] = \mathbb{E}[f(\boldsymbol{x})\mathcal{H}_{d,\ell}(\boldsymbol{x}^{\otimes \ell})] = \boldsymbol{A}_{d,\ell},$$

从而 $\boldsymbol{A}_{d,\ell} \in \mathrm{TSym}_\ell^{\boldsymbol{W}}(\mathbb{R}^d)$，对所有 $\ell \geq 0$ 均成立。由引理 18，存在张量 $\boldsymbol{C}_{d,\ell} \in \mathrm{Sym}_\ell(\mathbb{R}^d)$，使得

$$\boldsymbol{A}_{d,\ell} = \sum_{j=0}^{\lfloor \ell/2 \rfloor} \mathsf{P}_{\mathsf{sym}}(((\boldsymbol{W}\boldsymbol{W}^{\mathsf{T}})^{\otimes(\ell-2j)}\boldsymbol{C}_{d,\ell-2j}) \otimes (\mathbf{I}_d - \boldsymbol{W}\boldsymbol{W}^{\mathsf{T}})^{\otimes j}).$$

将 $(\mathbf{I}_d - \boldsymbol{W}\boldsymbol{W}^{\mathsf{T}})^{\otimes j}$ 展开为形如 $\boldsymbol{W}\boldsymbol{W}^{\mathsf{T}}$ 和 $\mathbf{I}_d$ 的分量的乘积，我们可以定义新的张量 $\tilde{\boldsymbol{C}}_{d,\ell-2m} \in \mathrm{Sym}_{\ell-2m}(\mathbb{R}^d)$、$m = 0, \ldots\ldots, \lfloor \ell/2 \rfloor$，使得

$$\boldsymbol{A}_{d,\ell} = \sum_{j=0}^{\lfloor \ell/2 \rfloor} \mathsf{P}_{\mathsf{sym}}(((\boldsymbol{W}\boldsymbol{W}^{\mathsf{T}})^{\otimes(\ell-2j)}\tilde{\boldsymbol{C}}_{d,\ell-2j}) \otimes \mathbf{I}_d^{\otimes j}).$$

与任意 $\boldsymbol{x} \in \mathbb{S}^{d-1}$ 的 $\mathcal{H}_{d,\ell}(\boldsymbol{x}^{\otimes \ell})$ 取内积，并利用 $\mathcal{H}_{d,\ell}(\boldsymbol{x}^{\otimes \ell})$ 无迹这一事实，我们得到

$$\langle \boldsymbol{A}_{d,\ell}, \mathcal{H}_{d,\ell}(\boldsymbol{x}^{\otimes \ell})\rangle_{\mathsf{F}} = \langle (\boldsymbol{W}\boldsymbol{W}^{\mathsf{T}})^{\otimes \ell}\tilde{\boldsymbol{C}}_\ell, \mathcal{H}_{d,\ell}(\boldsymbol{x}^{\otimes \ell})\rangle_{\mathsf{F}}.$$

由于这对所有 $\ell \geq 0$ 都成立，

$$f(\boldsymbol{x}) = \sum_{\ell \geq 0} \langle \mathsf{P}_{\mathsf{tf}}((\boldsymbol{W}\boldsymbol{W}^{\mathsf{T}})^{\otimes \ell}\tilde{\boldsymbol{C}}_\ell), \mathcal{H}_{d,\ell}(\boldsymbol{x}^{\otimes \ell})\rangle_{\mathsf{F}}$$

其中，由上述可知，$\mathsf{P}_{\mathsf{tf}}((\boldsymbol{W}\boldsymbol{W}^{\mathsf{T}})^{\otimes \ell}\tilde{\boldsymbol{C}}_\ell) = \boldsymbol{A}_{d,\ell}$。特别地，该引理的最后一部分在 $\boldsymbol{B}_{\mathsf{s},\ell} = (\boldsymbol{W}^{\mathsf{T}})^{\otimes \ell}\tilde{\boldsymbol{C}}_{d,\ell} \in \mathrm{Sym}_\ell(\mathbb{R}^{\mathsf{s}})$ 时成立。 □

引理 20 表明任意 $\mathcal{O}_d^{\boldsymbol{W}}$-不变函数 $f \in L^2(\mathbb{S}^{d-1}, \tau_d)$ 可展开为

$$f(\boldsymbol{x}) = \sum_{\ell \geq 0} \langle \mathsf{P}_{\mathsf{tf}}(\boldsymbol{W}^{\otimes \ell}\boldsymbol{B}_{\mathsf{s},\ell}), \mathcal{H}_{d,\ell}(\boldsymbol{x}^{\otimes \ell})\rangle_{\mathsf{F}}$$

对某些 $\boldsymbol{B}_{\mathsf{s},\ell} \in \mathrm{Sym}_\ell(\mathbb{R}^{\mathsf{s}})$。由于张量 $\{\boldsymbol{B}_{\mathsf{s},\ell}\}_{\ell \geq 0}$ 是有限维的（维数不依赖于 d），该展开更便于在高维极限 $d \to \infty$ 下进行分析。例如，我们可以证明，当 d 很大时，$\mathsf{P}_{\mathsf{tf}}(\boldsymbol{W}^{\otimes \ell}\boldsymbol{B}_{\mathsf{s},\ell})$ 可由 $\boldsymbol{W}^{\otimes \ell}\boldsymbol{B}_{\mathsf{s},\ell}$ 很好地逼近。

引理 21. 设 d、s $\ell \in \mathbb{N}$ 与 $\boldsymbol{W} \in \mathbb{R}^{d\times \mathsf{s}}$ 为一个标准正交 s-标架。若 $d \geq 4\ell^4\mathsf{s}$，则对任意 $\boldsymbol{B}_{\mathsf{s},\ell} \in \mathrm{Sym}_\ell(\mathbb{R}^{\mathsf{s}})$，

$$\left\| \mathsf{P}_{\mathsf{tf}}\left(\boldsymbol{W}^{\otimes \ell}\boldsymbol{B}_{\mathsf{s},\ell}\right) - \boldsymbol{W}^{\otimes \ell}\boldsymbol{B}_{\mathsf{s},\ell} \right\|_{\mathsf{F}} \leq \left\| \mathsf{P}_{\mathsf{tf}}\left(\boldsymbol{W}^{\otimes \ell}\boldsymbol{B}_{\mathsf{s},\ell}\right) \right\|_{\mathsf{F}} \frac{2\ell^2\sqrt{\mathsf{s}}}{\sqrt{d}}.$$

证明。 根据（86），

$$\mathsf{P}_{\mathsf{tf}}\left(\boldsymbol{W}^{\otimes \ell}\boldsymbol{B}_{\mathsf{s},\ell}\right) - \boldsymbol{W}^{\otimes \ell}\boldsymbol{B}_{\mathsf{s},\ell} = \sum_{j=1}^{\lfloor \ell/2 \rfloor} h_{\ell,j}^{(d)} \mathsf{P}_{\mathsf{sym}}\left(\tau^j\left(\boldsymbol{W}^{\otimes \ell}\boldsymbol{B}_{\mathsf{s},\ell}\right) \otimes \mathbf{I}_d^{\otimes j}\right)$$

其中，系数 $h_{\ell,j}^{(d)}$ 在（88）中有明确的定义。取弗罗贝尼乌斯范数，并利用三角不等式，以及 $\mathsf{P}_{\mathsf{sym}}$ 作为正交投影具有非扩张性这一事实，

$$\left\|\mathsf{P}_{\mathsf{tf}}\left(\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell}\right)-\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell}\right\|_{\mathsf{F}}\le\sum_{j=1}^{\lfloor\ell/2\rfloor}|h_{\ell,j}^{(d)}|\|\tau^{j}(\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell})\otimes\mathbf{I}_{d}^{\otimes j}\|_{\mathsf{F}}=\sum_{j=1}^{\lfloor\ell/2\rfloor}|h_{\ell,j}^{(d)}|d^{j/2}\|\tau^{j}(\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell})\|_{\mathsf{F}}.$$

根据偏迹算子的定义 τ，由式（15）可知

$$\tau\left(\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell}\right)=\boldsymbol{W}^{\otimes(\ell-2)}\tau(\boldsymbol{B}_{\mathsf{s},\ell}).$$

将此过程迭代 j 次，得到

$$\tau^{j}\left(\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell}\right)=\boldsymbol{W}^{\otimes(\ell-2j)}\tau^{j}(\boldsymbol{B}_{\mathsf{s},\ell})$$

对所有 $j=1$、……、$\lfloor\ell/2\rfloor$均成立。利用粗略估计 $\|\tau^{j}(\boldsymbol{B}_{\mathsf{s},\ell})\|_{\mathsf{F}}\le s^{j/2}\|\boldsymbol{B}_{\mathsf{s},\ell}\|_{\mathsf{F}}$，

$$\left\|\mathsf{P}_{\mathsf{tf}}\left(\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell}\right)-\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell}\right\|_{\mathsf{F}}\le\|\boldsymbol{B}_{\mathsf{s},\ell}\|_{\mathsf{F}}\sum_{j=1}^{\lfloor\ell/2\rfloor}|h_{\ell,j}^{(d)}|(d\mathsf{s})^{j/2}.$$

利用式（88）中系数 $h_{\ell,j}^{(d)}$ 的显式表达式，可得 $|h_{\ell,j}^{(d)}|\le\left(\ell^{2}/(2d)\right)^{j}$ 对所有 $j=1$、……、$\lfloor\ell/2\rfloor$均成立，因此

$$\left\|\mathsf{P}_{\mathsf{tf}}\left(\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell}\right)-\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell}\right\|_{\mathsf{F}}\le\|\boldsymbol{B}_{\mathsf{s},\ell}\|_{\mathsf{F}}\sum_{j=1}^{\lfloor\ell/2\rfloor}\left(\frac{\ell^{2}\sqrt{\mathsf{s}}}{2\sqrt{d}}\right)^{j}\tag{95}$$

当 $d\ge4\ell^{4}\mathsf{s}$ 时，右侧的因子满足

$$\sum_{j=1}^{\lfloor\ell/2\rfloor}\left(\frac{\ell^{2}\sqrt{\mathsf{s}}}{2\sqrt{d}}\right)^{j}\le\frac{\ell^{2}\sqrt{\mathsf{s}}}{\sqrt{d}}\le2^{-1}.$$

因此，由式（95）及三角不等式可得，

$$\frac{1}{2}\|\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell}\|_{\mathsf{F}}\le\|\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell}\|_{\mathsf{F}}-\left\|\mathsf{P}_{\mathsf{tf}}\left(\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell}\right)-\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell}\right\|_{\mathsf{F}}\le\|\mathsf{P}_{\mathsf{tf}}(\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell})\|_{\mathsf{F}},$$

这意味着 $\|\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell}\|_{\mathsf{F}}\le2\|\mathsf{P}_{\mathsf{tf}}(\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell})\|_{\mathsf{F}}$。 □

在我们关注的 $d\to\infty$ 与 s 固定 ℓ 情形下，引理 21 表明，$\mathsf{P}_{\mathsf{tf}}(\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell})$ 可以很好地近似为 $\boldsymbol{W}^{\otimes\ell}\boldsymbol{B}_{\mathsf{s},\ell}$，其乘法误差率随 $d\to\infty$ 趋于零。

D.4 张量积表示

我们考虑张量积表示 $\mathrm{TSym}^{p}(\mathbb{R}^{d})\otimes\mathrm{TSym}^{q}(\mathbb{R}^{d})$ 在对角作用 $g\cdot(\boldsymbol{A}\otimes\boldsymbol{B})=(g\cdot\boldsymbol{A})\otimes(g\cdot\boldsymbol{B})$ 下，它具有式 (25) 给出的不可约分解。下一个引理给出张量 $\boldsymbol{A}\otimes\boldsymbol{B}$ 的显式分解。一般情形 $\boldsymbol{C}\in\mathrm{TSym}^{p}(\mathbb{R}^{d})\otimes\mathrm{TSym}^{q}(\mathbb{R}^{d})$ 可由线性性得出。

引理 22。 设 $p, q \geq 0$。那么，对所有 $\boldsymbol{A} \in \mathrm{TSym}^p(\mathbb{R}^d)$ 和 $\boldsymbol{B} \in \mathrm{TSym}^q(\mathbb{R}^d)$，有

$$\langle \boldsymbol{A} \otimes \boldsymbol{B}, \mathcal{H}_{d,p}(\boldsymbol{z}) \otimes \mathcal{H}_{d,q}(\boldsymbol{z}) \rangle_{\mathsf{F}} = \sum_{j=0}^{p \wedge q} b_{p,q,j}^{(d)} \langle \boldsymbol{A} \diamond_j \boldsymbol{B}, \mathcal{H}_{d,p+q-2j}(\boldsymbol{z}) \rangle_{\mathsf{F}},$$

其中 $\diamond_j$ 表示在 (26)中定义的双线性算子，而标量 $\{b_{p,q,j}^{(d)}\}_{j=0}^{p \wedge q}$ 定义为

$$b_{p,q,j}^{(d)} = \frac{f_{p+q,j}^{(d)} \kappa_{d,p} \kappa_{d,q} \sqrt{N_{d,p} N_{d,q}}}{\kappa_{d,p+q-2j} \sqrt{N_{d,p+q-2j}}} \frac{2^j (p!)(q!)(p+q-2j)!}{(p+q)!(p-j)!(q-j)!}$$

证明。 由 (91) 和 (94),

$$\begin{aligned} \langle \mathcal{H}_{d,p}(\boldsymbol{z}) \otimes \mathcal{H}_{d,q}(\boldsymbol{z}), \boldsymbol{A} \otimes \boldsymbol{B} \rangle_{\mathsf{F}} &= \kappa_{d,p} \kappa_{d,q} \sqrt{N_{d,p} N_{d,q}} \langle \boldsymbol{z}^{\otimes (p+q)}, \boldsymbol{A} \otimes \boldsymbol{B} \rangle_{\mathsf{F}} \\ &= \sum_{j=0}^{\lfloor (p+q)/2 \rfloor} \frac{f_{p+q,j}^{(d)} \kappa_{d,p} \kappa_{d,q} \sqrt{N_{d,p} N_{d,q}}}{\kappa_{d,p+q-2j} \sqrt{N_{d,p+q-2j}}} \left\langle \mathsf{P}_{\mathsf{sym}} \left(\mathcal{H}_{d,p+q-2j}(\boldsymbol{z}) \otimes \mathbf{I}_d^{\otimes j} \right), \boldsymbol{A} \otimes \boldsymbol{B} \right\rangle_{\mathsf{F}}. \end{aligned}$$

暂且固定 $j \in [\lfloor (p+q)/2 \rfloor]$，并考虑项 $\langle \mathsf{P}_{\mathsf{sym}}(\mathcal{H}_{d,\ p+q-2j}(\boldsymbol{z}) \otimes \mathbf{I}_d^{\otimes j})\ \boldsymbol{A} \otimes \boldsymbol{B} \rangle_{\mathsf{F}}$。由于 $\mathsf{P}_{\mathsf{sym}}$ 是正交投影，我们可以将其移至第二个自变量，得到

$$\langle \mathsf{P}_{\mathsf{sym}}(\mathcal{H}_{d,p+q-2j}(\boldsymbol{z}) \otimes \mathbf{I}_d^{\otimes j}), \boldsymbol{A} \otimes \boldsymbol{B} \rangle_{\mathsf{F}} = \langle \mathcal{H}_{d,p+q-2j}(\boldsymbol{z}), \mathsf{P}_{\mathsf{sym}}(\boldsymbol{A} \otimes \boldsymbol{B})[\mathbf{I}_d^{\otimes j}] \rangle_{\mathsf{F}}$$

其中，$\mathsf{P}_{\mathsf{sym}}(\boldsymbol{A} \otimes \boldsymbol{B})[\mathbf{I}_d^{\otimes j}] \in \mathrm{Sym}^{p+q-2j}(\mathbb{R}^d)$ 表示沿 $\mathsf{P}_{\mathsf{sym}}(\boldsymbol{A} \otimes \boldsymbol{B})$ 最后 $2j$ 个分量进行缩并。令 $G \subseteq \mathfrak{S}_{p+q}$ 表示交叉配对置换的集合，即满足对所有 $k \in \{p+q-2j+1, \ldots, p+q-j\}$，均有 $|\{\pi(k)\ \pi(k+j)\} \cap [p]| = 1$ 的置换 $\pi \in \mathfrak{S}_{p+q}$ 的集合。若 $\pi \notin G$，则存在某个 $k \in \{p+q-2j+1, \ldots, p+q-j\}$，使得 $\pi(k)\ \pi(k+j) \leq p$ 或 $\pi(k)\ \pi(k+j) > p$ 两者之一成立。对指标 $i_{p+q-2j+1}, \ldots, i_{p+q}$ 求和后结果为零，因为 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 都是无迹的。因此，

$$\mathsf{P}_{\mathsf{sym}}(\boldsymbol{A} \otimes \boldsymbol{B})[\mathbf{I}_d^{\otimes j}] = \frac{1}{(p+q)!} \sum_{\pi \in G} (\boldsymbol{A} \otimes \boldsymbol{B})^{\pi}[\mathbf{I}_d^{\otimes j}].$$

每个 $\pi \in G$ 都在 $\boldsymbol{A}$ 的 j 个槽位与 $\boldsymbol{B}$ 的 j 个槽位之间诱导出一个匹配。由于 $\boldsymbol{A}$，$\boldsymbol{B}$ 是对称张量，我们可以分别对 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的槽位进行置换，而不改变求和项的值。特别地，存在一个置换 $\sigma_\pi \in \mathfrak{S}_{p+q-2j}$，使得

$$(\boldsymbol{A} \otimes \boldsymbol{B})^{\pi}[\mathbf{I}_d^{\otimes j}] = (\boldsymbol{A} \otimes_j \boldsymbol{B})^{\sigma_\pi},$$

因此，

$$\mathsf{P}_{\mathsf{sym}} \left(\mathsf{P}_{\mathsf{sym}}(\boldsymbol{A} \otimes \boldsymbol{B})[\mathbf{I}_d^{\otimes j}] \right) = \frac{|G|}{(p+q)!} \mathsf{P}_{\mathsf{sym}}(\boldsymbol{A} \otimes_j \boldsymbol{B}).$$

剩下的 是计算 $|G|$，这通过组合论论证来完成，以得出 in

$$|G| = \frac{(p!)(q!)2^j (p+q-2j)!}{(p-j)!(q-j)!}$$

当 $j \leq \min(p, q)$，且 $|G| = 0$ 否则。结合上述观察以及式（ $\diamond_j$ （26）中双线性算子的定义，证明得证。 □

D.5 矩阵系数的超收缩性

下面，我们证明正交群 $\mathcal{O}_d$ 作用在单位球面上的函数中，次数为 k 的矩阵系数所构成的子空间具有超收缩性。证明依赖于连通分支 $\mathcal{SO}_d$ 上热半群的经典对数 Sobolev 结果和超收缩结果；为完整起见，我们在此简要回顾这些结果。

正交群存在半直积分解 $\mathcal{O}_d = \mathcal{SO}_d \rtimes \mathcal{R}$，其中 $\mathcal{R} = \{e, \mathrm{R}\}$，R 是任意固定的反射（即 $\mathrm{R}^2 = e$ 且 $\det \mathrm{R} = -1$，其中 e 是单位元）。由于 $\mathcal{SO}_d$ 是 $\mathcal{O}_d$ 的正规子群，$\mathcal{O}_d$ 在 $L^2(\mathbb{S}^{d-1})$ 上的作用表示限制为 $\mathcal{SO}_d$ 在 $L^2(\mathbb{S}^{d-1})$ 上的作用表示，并且具有相同的球谐子空间半单分解。特别地，每个球谐子空间 $\mathscr{S}_{d,\ell}$ 都是 $\mathcal{O}_d$ 和 $\mathcal{SO}_d$ 的不可约表示。

固定 $\ell \in \mathbb{N}$。对于 $f, h \in \mathscr{S}_{d,\ell}$ 和 ρ 在 $L^2(\mathbb{S}^{d-1})$ 上的作用表示，定义矩阵系数

$$\rho_{fh}(g) := \langle \rho(g) \cdot f, h \rangle_{L^2}, \qquad g \in \mathcal{O}_d,$$

我们可以将其视为 $L^2(\mathcal{SO}_d)$ 或 $L^2(\mathcal{O}_d)$ 中的元素。令

$$\mathcal{M}_{d,\ell} = \operatorname{span}\{\rho_{fh} : f, h \in \mathscr{S}_{d,\ell}\}$$

设为 $L^2(\mathcal{SO}_d)$ 中与不可约表示 $\mathscr{S}_{d,\ell}$ 相关联的等型子空间。

令 Δ 表示 $\mathcal{SO}_d$ 上关于标准双不变黎曼度量的拉普拉斯–贝尔特拉米算子。由于 $\mathcal{SO}_d$ 是一个紧李群，拉普拉斯–贝尔特拉米算子与左正则表示中作用的卡西米尔算子重合（例如，参见 [Far08，第 12]章）。因此，Δ 在 $L^2(\mathcal{SO}_d)$ 的每个等型分量上均按标量作用 [Far08，推论 6.7。2], [GW09，引理 3.3。8]。特别地，每个$F \in \mathcal{M}_d$，ℓ 都是 Δ 的特征函数，其特征值等于$\mathscr{S}_d$的卡西米尔特征值的相反数，ℓ [Far08，命题 8.2。1]。由于 $\mathscr{S}_d$，ℓ的最高权为 $\lambda = (\ell, 0, \ldots, 0)$，由卡西米尔特征值的标准公式 [Far08，命题 12.1。2] 可得

$$\Delta F = -\lambda_{d,\ell} F, \qquad F \in \mathcal{M}_{d,\ell}, \qquad \lambda_{d,\ell} = \ell(\ell + d - 2), \tag{96}$$

令 $(P_t)_{t \geq 0} = (e^{t\Delta})_{t \geq 0}$ 表示相应的热半群。一个经典结果表明，$\mathcal{SO}_d$满足常数为 $c_{\mathrm{LS}}(\mathcal{SO}_d) = \frac{4}{d-2}$ 的对数 Sobolev 不等式（LSI），其中 $d \geq 3$。这可由 Bakry–Émery 曲率–维数判据推出 [BGL13, Theorem 6.8.1]。事实上，相对于由 Killing 型的负值诱导的度量，$\mathcal{SO}_d$的里奇曲率下界为 $1/4$ [Rot86, Theorem 3]。由于 Killing 度量与 $\mathcal{SO}_d$ 上的标准双不变度量仅相差一个常数缩放因子 $2(d-2)$，因此根据对数 Sobolev 不等式的缩放性质，可得所述 LSI 常数。由 Gross 定理 [BGL13, Theorem 5.2.3], ，这意味着 (P_t)具有超收缩性：对于 $1 < p \leq q < \infty$ 和 $t \geq 0$,

$$\|P_t F\|_{L^q(\mathcal{SO}_d)} \leq \|F\|_{L^p(\mathcal{SO}_d)} \quad \text{whenever} \quad q - 1 \leq e^{2t/c_{\mathrm{LS}}(\mathcal{SO}_d)}(p-1). \tag{97}$$

结合 (96) 和 (97) 可得作用于 $\mathcal{SO}(d)$ 的超收缩不等式，对每个次数为 ℓ 的块 :

$$\|F\|_{L^q(\mathcal{SO}_d)} \leq \left(\tfrac{q-1}{p-1}\right)^{\gamma_d(\ell)} \|F\|_{L^p(\mathcal{SO}_d)}, \qquad \gamma_d(\ell) = \frac{2\ell(\ell+d-2)}{d-2}. \tag{98}$$

$\mathcal{O}_d$ 的超收缩结果可通过写出 $\mathcal{O}_d = \mathcal{SO}_d \sqcup \mathrm{R}\mathcal{SO}_d$，并对两个分量应用上述超收缩性得出。

引理 23（矩阵系数的超收缩性）。设 $d \geq 3$ 和 $\ell \geq 1$。对于所有 $F \in \mathcal{M}_{d,\ell}$ 和 $1 < p \leq q < \infty$

$$\|F\|_{L^q(\mathcal{O}_d)} \leq 2^{\frac{1}{p}-\frac{1}{q}} \left(\frac{q-1}{p-1}\right)^{\gamma_d(\ell)} \|F\|_{L^p(\mathcal{O}_d)}, \qquad \gamma_d(\ell) = \frac{2\ell(\ell+d-2)}{d-2}.$$

证明。令 $\mathsf{R} \in \mathcal{O}_d$ 为任意固定的反射，并将 $\mathcal{O}_d$ 分解为 $\mathcal{O}_d = \mathcal{SO}_d \sqcup \mathcal{SO}_d\mathsf{R}$。对于 $F \in \mathcal{M}_{d,\ell}$注意到 $F \circ \mathsf{R} \in \mathcal{M}_{d,\ell}$ 并且也满足（98）。因此，

$$\begin{aligned}
\|F\|_{L^q(\mathcal{O}_d)}^q &= \frac{1}{2}\|F\|_{L^q(\mathcal{SO}_d)}^q + \frac{1}{2}\|F \circ \mathsf{R}\|_{L^q(\mathcal{SO}_d)}^q \\
&\leq \frac{1}{2}\left(\frac{q-1}{p-1}\right)^{q\gamma_d(\ell)} \left\{\|F\|_{L^p(\mathcal{SO}_d)}^q + \|F \circ \mathsf{R}\|_{L^p(\mathcal{SO}_d)}^q\right\} \\
&\leq \frac{1}{2}\left(\frac{q-1}{p-1}\right)^{q\gamma_d(\ell)} \left\{\|F\|_{L^p(\mathcal{SO}_d)}^p + \|F \circ \mathsf{R}\|_{L^p(\mathcal{SO}_d)}^p\right\}^{q/p} = 2^{\frac{q}{p}-1}\left(\frac{q-1}{p-1}\right)^{q\gamma_d(\ell)} \|F\|_{L^p(\mathcal{O}_d)}^q,
\end{aligned}$$

这里我们用到了以下事实：$a^r + b^r \leq (a+b)^r$ 对 $a,\ b \geq 0$ 和 $r \geq 1$成立。取 q 次方根即可完成证明。 □

D.6 矩阵集中不等式

下面，我们给出一个关于独立随机矩阵之和的矩阵集中不等式，该不等式源自一系列研究工作 [BvH24，BBvH23]。类似版本已用于相近的场景，例如参见 [DLB25，DPVLB24，JKMS25]。

给定一个随机矩阵 $\boldsymbol{X} = \sum_{i\in[n]} \boldsymbol{X}_i \in \mathbb{R}^{p\times q}$，定义

$$\begin{aligned}
\sigma(\boldsymbol{X})^2 &= \max\left\{\|\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\mathsf{T}]\|_{\mathrm{op}}, \|\mathbb{E}[\boldsymbol{X}^\mathsf{T}\boldsymbol{X}]\|_{\mathrm{op}}\right\}, \\
\sigma_*(\boldsymbol{X})^2 = \sup_{\|\boldsymbol{u}\|_2=\|\boldsymbol{v}\|_2=1} \mathbb{E}[\langle \boldsymbol{u}, \boldsymbol{X}\boldsymbol{v}\rangle^2], &\quad \bar{R}(\boldsymbol{X})^2 = \mathbb{E}\left[\max_{1\leq i\leq p} \|\boldsymbol{X}_i\|_{\mathrm{op}}^2\right], \\
v(\boldsymbol{X})^2 &= \|\mathsf{Cov}(\boldsymbol{X})\|_{\mathrm{op}}.
\end{aligned} \tag{99}$$

这里，$\|\mathsf{Cov}(\boldsymbol{X})\|_{\mathrm{op}}$ 表示随机矩阵 $\boldsymbol{X}$作为 $\mathbb{R}^{pq}$中的向量时，其协方差算子的算子范数。注意，

$$\begin{aligned}
\|\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\mathsf{T}]\|_{\mathrm{op}} &= \sup_{\|\boldsymbol{u}\|_2=1} \mathbb{E}[\|\boldsymbol{X}^\mathsf{T}\boldsymbol{u}\|_2^2] \\
&= \sup_{\|\boldsymbol{u}\|_2=1} \sum_{j=1}^{p} \mathbb{E}[\langle \boldsymbol{e}_j, \boldsymbol{X}^\mathsf{T}\boldsymbol{u}\rangle^2] \leq p\sigma_*(\boldsymbol{X})^2.
\end{aligned}$$

类似地，$\|\mathbb{E}[\boldsymbol{X}^\mathsf{T}\boldsymbol{X}]\|_{\mathrm{op}} \leq q\sigma_*(\boldsymbol{X})^2$ 使得

$$\sigma(\boldsymbol{X})^2 \leq (p \vee q)\sigma_*(\boldsymbol{X})^2. \tag{100}$$

引理 24 (矩阵集中不等式)。设 $\{\boldsymbol{X}_i\}_{i=1}^n \subseteq \mathbb{R}^{p\times q}$ 为相互独立、均值为零的随机矩阵，且 $\boldsymbol{X} = \sum_{i=1}^n \boldsymbol{X}_i$。那么，存在一个通用常数 $C > 0$，使得对于所有 $t > 0$ 和 $R \geq \bar{R}(\boldsymbol{X})^{1/2}\sigma(\boldsymbol{X})^{1/2} + 2^{1/2}\bar{R}(\boldsymbol{X})$，满足 $\mathbb{P}(\max_{i\in[n]} \|\boldsymbol{X}_i\|_{\mathrm{op}} > R) \leq \delta$，

$$\|\boldsymbol{X}\|_{\mathrm{op}} \leq C\left(2\sigma(\boldsymbol{X}) + \sigma(\boldsymbol{X})^{1/2}v(\boldsymbol{X})^{1/2}\log(p+q)^{3/4} + \sigma_*(\boldsymbol{X})t^{1/2} + R^{1/3}\sigma(\boldsymbol{X})^{2/3}t^{2/3} + Rt\right),$$

以至少为 $1 - \delta - (p+q+1)\,e^{-t}$ 的概率。

证明。 对于每个 $i \in [n]$，将 $\boldsymbol{X}_i$ 的对称扩张定义为

$$\check{\boldsymbol{X}}_i = \begin{bmatrix} \mathbf{0} & \boldsymbol{X}_i \\ \boldsymbol{X}_i^\mathsf{T} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(p+q)\times(p+q)}.$$

确实，$\{\check{\boldsymbol{X}}_i\}_{i=1}^n$ 是独立自伴随机矩阵。根据 [BBvH23，引理 4。10]，$\sigma(\boldsymbol{X}) = \sigma(\check{\boldsymbol{X}})$ 和 $\sigma_*(\boldsymbol{X}) = \sigma_*(\check{\boldsymbol{X}})$。此外，

$$\|\check{\boldsymbol{X}}_i\|_{\mathrm{op}}^2 = \|\check{\boldsymbol{X}}_i^2\|_{\mathrm{op}} = \max\{\|\boldsymbol{X}_i\boldsymbol{X}_i^\mathsf{T}\|_{\mathrm{op}}, \|\boldsymbol{X}_i^\mathsf{T}\boldsymbol{X}_i\|_{\mathrm{op}}\} = \|\boldsymbol{X}_i\|_{\mathrm{op}}^2,$$

因此 $\bar{R}(\check{\boldsymbol{X}}) = \bar{R}(\boldsymbol{X})$。根据 [BvH24，定理2。8]，以及 $\|\check{\boldsymbol{X}}\|_{\mathrm{op}} = \|\boldsymbol{X}\|_{\mathrm{op}}$这一事实，存在一个通用常数 $C > 0$ 使得：

$$\mathbb{P}\left(\|\boldsymbol{X}\|_{\mathrm{op}} - \|\boldsymbol{G}\|_{\mathrm{op}} \geq C\epsilon_R(t), \max_{1\leq i\leq n}\|\boldsymbol{X}_i\|_{\mathrm{op}} \leq R\right) \leq (p+q)e^{-t},$$

对所有 $t > 0$ 和 $R \geq \bar{R}(\boldsymbol{X})^{1/2}\sigma(\boldsymbol{X})^{1/2} + 2^{1/2}\bar{R}(\boldsymbol{X})$，其中

$$\epsilon_R(t) = \sigma_*(\boldsymbol{X})t^{1/2} + R^{1/3}\sigma(\boldsymbol{X})^{2/3}t^{2/3} + Rt,$$

以及 $\boldsymbol{G} = \sum_{i=1}^n \boldsymbol{G}_i$，其中 $\{\boldsymbol{G}_i\}_{i=1}^n$ 为满足 $\mathbb{E}[\boldsymbol{G}_i \otimes \boldsymbol{G}_i] = \mathbb{E}[\boldsymbol{X}_i \otimes \boldsymbol{X}_i]$ 的独立中心高斯随机矩阵。

接下来，根据 [BBvH23、推论 2.2 和引理 2。5]，存在一个通用常数 $C' > 0$ 使得，对所有 $t > 0$，

$$\mathbb{P}\left(\|\boldsymbol{G}\|_{\mathrm{op}} \geq 2\sigma(\boldsymbol{X}) + C'\sigma(\boldsymbol{X})^{1/2}\|\mathsf{Cov}(\boldsymbol{X})\|^{1/4}\log(p+q)^{3/4} + C'\sigma_*(\boldsymbol{X})t\right) \leq e^{-t^2},$$

其中，$\mathsf{Cov}(\boldsymbol{X})$ 是将随机矩阵 $\boldsymbol{X}$ 视为 $\mathbb{R}^{pq}$ 中的向量时的协方差。结合上述两个不等式可得，存在一个通用常数 $C'' > 0$，使得对所有$t > 0$ 和 $R \geq \bar{R}(\boldsymbol{X})^{1/2}\sigma(\boldsymbol{X})^{1/2} + 2^{1/2}\bar{R}(\boldsymbol{X})$，

$$\mathbb{P}\left(\max_{i\in[n]}\|\boldsymbol{X}_i\|_{\mathrm{op}} > R\right) \leq \delta,$$

then

$$\|\boldsymbol{X}\|_{\mathrm{op}} \leq C''\left(\sigma_*(\boldsymbol{X})t^{1/2} + R^{1/3}\sigma(\boldsymbol{X})^{2/3}t^{2/3} + Rt + \sigma(\boldsymbol{X})^{1/2}\|\mathsf{Cov}(\boldsymbol{X})\|^{1/4}\log(p+q)^{3/4} + 2\sigma(\boldsymbol{X})\right),$$

其概率至少为 $1 - (p+q+1)e^{-t} - \delta$。 □