arXiv:2302.07216v1 [stat.ME] 14 Feb 2023

论多路主成分分析*

Jialin Ouyang 和 Ming Yuan

统计学系

哥伦比亚大学

（2023年2月15日）

摘要

多路数据正变得越来越普遍。尽管已有许多方法可将主成分分析（PCA）从传统数据矩阵推广到多维数组，但这些方法与传统主成分分析在概念上的差异，以及此类差异所带来的方法论影响，仍在很大程度上尚不明确。本文旨在专门探讨这些问题。具体而言，我们阐明了多路数据中主成分分析与奇异值分解（SVD）之间的细微差异，并表明，即使不存在传统主成分分析所要求的特征值间隙，多路主成分（PCs）仍可得到可靠估计，而且通常比传统主成分更高效。此外，样本多路主成分在渐近意义下相互独立，因此能够分别对总体主成分进行更准确的推断。最后，通过数值实例（包括模拟数据和真实数据），进一步展示了多路主成分分析的实际优点。

1 引言

在实际应用中，我们越来越常需要处理结构丰富而复杂的数据。这类数据更适合组织成多维数组，而非通常的数据矩阵。这种多路数据在化学计量学等众多领域中随处可见，

*本研究部分得到NSF资助项目DMS-2015285和DMS-2052955的支持。

经济学、心理测量学和信号处理等领域（例如，Kroonenberg，2008年；Anandkumar 等，2014；Zhang 和 Xia，2018；Chen 等，2020a、b；Han 等，2020；Xia 等，2020；Bi 等，2021；Chen 等，2021；Han 等，2022）。本文探讨主成分分析（PCA）应用于这类数据的方法论意义和统计性质，并指出这样做的优势与挑战。

当数据以矩阵形式组织时，PCA 是多元数据分析中最常用的统计方法之一。例如，参见 Anderson (1984);Jolliffe (2002)。将数据矩阵的每个列向量视为一个观测，PCA 寻找这些向量到新坐标系的正交线性变换，使得各坐标的方差依次达到最大。借助这种方法，我们可以用少量坐标表示数据中的大部分变异，从而为降低维数提供指导。因此，PCA 通常作为许多后续分析中捕捉数据集基本特征的关键第一步，并广泛应用于众多科学与工程领域。从矩阵推广到多路数据时，每个观测本身构成一个矩阵，或更一般地说，构成一个多维数组。例如，当在不同地点和时间组合下进行重复测量时，每个观测更自然地可以组织为一个矩阵，其中每一行对应某个地点，每一列对应某个时间点。要将 PCA 应用于这类数据，一个看似可行的做法是忽略观测的多路性质，仍将每个观测视为一个向量，这种做法通常称为串联。然而，正如众多实际应用所表明的，在应用 PCA 时适当考虑额外的结构，可以显著增强可解释性并提高效率。例如，参见Kroonenberg (2008年)。

针对这一目的，发展合适方法的历史悠久而辉煌，至少可以追溯到 20 世纪 60 年代 Tucker、Harshman 和 Carroll 的开创性工作。此后，人们还开发了许多方法，例如 Kroonenberg 和 De Leeuw（1980）；De Lathauwer 等（2000）；Vasilescu 和 Terzopoulos（2002）；Yang 等（2004）；Kong 等（2005）；Zhang 和 Zhou（2005）；Lu 等（2006，2008）；Li 等（2010）；Liu 等（2017）；Taguchi（2018），等等。关于现有技术的最新综述，可参见 Lu 等（2011）和 Cichocki 等（2015）。这些进展大多不属于统计学主流文献，而且通常带有浓厚的算法特征，并以探索性数据分析为重点。这些方法直观易懂，

与先将数据串联后再直接应用主成分分析相比，它们往往能得到更易于解释的见解。然而，这些方法的统计学基础在很大程度上尚不明确。本文的主要目标正是填补这一空白。事实上，正如下文将要展示的，审慎而严谨的统计学处理有助于更好地理解多路主成分分析的运行特性，改进方法，并揭示分析多路数据时面临的新机遇与挑战。

更具体地说，我们聚焦于一种简单而自然的多路主成分分析方法：在寻找使方差最大化的线性变换时，我们额外施加约束，要求这些变换符合数据的多路结构。这样做不仅提升了可解释性，也继承了传统主成分分析的许多优良性质。正如传统主成分（PC）是协方差矩阵的特征向量，多路主成分也可以由协方差算子的某些特征向量来识别。为了更好地理解多路结构对我们恢复多路主成分并据此进行推断的能力所产生的影响，我们还研究了尖峰协方差模型下多路主成分分析的性质。

在样本量较大而变量数量较少的经典情形下，传统主成分分析的统计性质已得到充分理解（例如，参见 Anderson，1984）。然而，在当今的应用中，维数也越来越常常很大。关于高维情形下传统主成分分析的理论结果也十分丰富，尤其是在尖峰协方差模型的背景下。例如，Johnstone 和 Lu（2009）首次通过证明：在信号强度固定时，当且仅当变量数量的阶低于样本量时，样本主成分分析才具有一致性，展示了维数在主成分分析中的关键作用。在另一篇有影响力的论文中，Paul（2007）建立了样本主成分的渐近分布。其他相关研究包括 Baik 和 Silverstein（2006）；Nadler（2008）；Johnstone 和 Lu（2009）；Jung 和 Marron（2009）；Bai 和 Silverstein（2010）；Lee 等（2010）；Benaych-Georges 和 Nadakuditi（2011）；Bai 和 Yao（2012）；Shen 等（2013）；Koltchinskii 和 Lounici（2014）；Koltchinskii 等（2017）；Wang 和 Fan（2017）；Koltchinskii 等（2020）等，不一而足。从某种意义上说，我们的结果自然地将这些早期工作推广到了多路主成分分析。然而，与协方差矩阵不同，需要处理高阶协方差算子，这带来了全新且根本性的挑战，也促使我们采用不同的证明策略并开发若干

新的技术工具。更重要的是，我们的分析还揭示了传统主成分分析与多路主成分分析在行为上的根本差异，并为后者的方法学发展带来了启示。

首先，我们在温和的正则条件下确立了样本多路主成分的收敛速率。这些速率解释了为何在对多路数据应用主成分分析时，必须考虑其固有的数据结构，也说明了为何在串联后直接套用主成分分析可能会带来问题。直观而言，多路主成分分析所用的参数少于传统主成分分析，因此更容易估计。我们的结果对此作出了精确描述：多路主成分的估计误差取决于数据数组各个模态的维度，而非条目总数。因此，即使条目总数远远超过样本量，也能准确估计多路主成分。但更重要的一点是，多路主成分的可估计程度取决于协方差算子中相应的特征值，而不像传统主成分分析那样取决于特征值之间的间隔。这一多少有些出人意料的发现具有深远影响。尤其是，这意味着对于多路数据，即使相应的特征值不是单重的，也仍然能够很好地估计主成分。

此外，为便于对多路主成分进行统计推断，我们推导了样本多路主成分的渐近分布。我们的结果再次揭示了多路主成分分析与传统主成分分析之间出乎意料但重要的差异。例如，估计得到的多路主成分彼此渐近独立，其渐近分布由相应的特征值而非特征值间隙决定。此外，我们还表明，对样本多路主成分进行偏差校正十分重要。与传统主成分分析类似，当维度（即各模态的维度）较高时，样本多路主成分可能存在显著偏差。但在缺少特征值间隙时，主成分的排序本身存在固有歧义，也可能产生另一种偏差来源。尽管如此，我们证明这两种偏差都可以消除，从而能够对多路主成分进行推断并构造其置信区间。

本文其余部分安排如下。在第2节中，我们介绍总体层面上的多路主成分分析概念，并说明其在有限样本中的应用。第3节研究样本多路主成分的收敛速率。接着，我们在第4节探讨多路主成分分析的渐近分布，并说明如何进行有效的

推断。第5节通过模拟和真实数据的数值实验，进一步展示多路主成分分析及我们所提出方法的优点。最后，我们在第6节进行总结。受篇幅所限，所有证明均置于补充材料中。

2 多路主成分分析

多路主成分分析可以从传统主成分分析的角度来理解，只是在主成分上施加了额外的多路结构。设 $\mathscr{X} \in \mathbb{R}^{d_1 \times \ldots \times d_p}$ 为一个 p 阶随机数组。为简化起见，除非另有说明，以下假设 $\mathscr{X}$ 是中心化的，即 $\mathbb{E}\mathscr{X} = 0$。主成分分析的基本思想是寻找一个关于 $\mathscr{X}$ 的线性变换，使其方差最大：

$$\max_{\mathscr{W} \in \mathbb{R}^{d_1 \times \cdots \times d_p} : \|\mathscr{W}\|_{\mathrm{F}} = 1} \operatorname{var}(\langle \mathscr{X}, \mathscr{W} \rangle). \tag{1}$$

这里 $\|\mathscr{W}\|_{\mathrm{F}} = \langle \mathscr{W}, \mathscr{W} \rangle^{1/2}$ 以及

$$\langle \mathscr{X}, \mathscr{W} \rangle = \sum_{j_1=1}^{d_1} \cdots \sum_{j_p=1}^{d_p} x_{j_1,\ldots,j_p} w_{j_1,\ldots,j_p}.$$

记 $\mathscr{U}_1$ 为（1）的解。多路主成分分析的基本前提是，$\mathscr{U}_1$ 符合 $\mathscr{X}$ 所具有的多路结构，即它是一个秩一张量，并且可以表示为

$$\mathscr{U}_1 = \mathbf{u}_1^{(1)} \otimes \mathbf{u}_1^{(2)} \otimes \cdots \otimes \mathbf{u}_1^{(p)}, \tag{2}$$

其中 $\mathbf{u}_1^{(q)}$ 是一个单位长度向量 $\mathbb{R}^{d_q}$，$\otimes$ 表示外积，即 $\mathscr{U}_1$ 的（i_1，……，i_p）个元素为

$$[\mathscr{U}_1]_{i_1,\ldots,i_p} = u_{1i_1}^{(1)} u_{1i_2}^{(2)} \cdots u_{1i_p}^{(p)}.$$

换句话说，$\mathscr{U}_1$ 也是下式的解

$$\max_{\mathscr{W} \in \Theta} \operatorname{var}(\langle \mathscr{X}, \mathscr{W} \rangle), \tag{3}$$

其中，Θ 表示所有维度相容且长度为单位长度的秩一张量的集合，即：

$$\Theta = \{\mathscr{W} = \mathbf{w}^{(1)} \otimes \mathbf{w}^{(2)} \otimes \cdots \otimes \mathbf{w}^{(p)} : \mathbf{w}^{(q)} \in \mathbb{R}^{d_q}, \|\mathbf{w}^{(q)}\| = 1, \forall q = 1, \ldots, p\}.$$

即使式（1）的解并非如式（2）所述的严格秩一形式，在寻求方差最大化变换时施加此类约束仍可能有益

因为这能增强可解释性：额外的秩一约束使我们能够分离沿各个模态的影响，并有助于解答诸如“谁在何时对谁做了什么”这样的问题，而这类问题往往是多模态数据分析的核心。另请参见Kroonenberg（2008年），其中有进一步的讨论和许多启发性示例。

后续主成分可以依次定义：

$$\max_{\substack{\mathscr{W}\in\mathbb{R}^{d_1\times\cdots\times d_p}:\|\mathscr{W}\|_{\mathrm{F}}=1\\ \mathscr{W}\perp\mathscr{U}_l,l=1,\ldots,k-1}} \mathrm{var}(\langle\mathscr{X},\mathscr{W}\rangle). \tag{4}$$

与前文一样，我们将考虑解的秩为一的情形。定义主成分时，一个关键要求是，第 k 个主成分与所有其他主成分正交，即 $\mathscr{W}\perp\mathscr{U}_l$。在向量情形中，也就是 $p=1$，这仅意味着 $\mathsf{h}\mathscr{W},\mathscr{U}_l\mathsf{i}=0$。然而，在多路情形中，正交性有许多不同的定义。关于这个问题的详细讨论，参见例如 Kolda（2001）。每种定义都有其自身的细微之处和注意事项，并可能带来不同的统计含义。本文将重点讨论完全正交性的定义：两个秩为一的张量 $\mathscr{W}_1=\mathbf{w}_1^{(1)}\otimes\mathbf{w}_1^{(2)}\otimes\cdots\otimes\mathbf{w}_1^{(p)}$ 和 $\mathscr{W}_2=\mathbf{w}_2^{(1)}\otimes\mathbf{w}_2^{(2)}\otimes\cdots\otimes\mathbf{w}_2^{(p)}$ 完全正交，当且仅当对于所有 $q=1$、.....、p，$\mathsf{h}\mathbf{w}_1^{(q)},\mathbf{w}_2^{(q)}\mathsf{i}=(\mathbf{w}_1^{(q)})^\top\mathbf{w}_2^{(q)}=0$。更具体地说，第 k 个多路主成分，记为 $\mathscr{U}_k$，满足

$$\max_{\mathscr{W}\in\Theta:\mathscr{W}\perp_c\mathscr{U}_l,\forall l<k} \mathrm{var}(\langle\mathscr{X},\mathscr{W}\rangle), \tag{5}$$

其中 $\perp_c$ 表示完全正交性。

与传统主成分分析的情形一样，多路主成分也可以用 $\mathrm{vec}(\mathscr{X})$ 的协方差矩阵等价地定义。事实上，对于多路数据，考虑协方差算子更为方便。具体而言，我们将视

$$\Sigma:=\mathrm{cov}(\mathscr{X})=\mathbb{E}(\mathscr{X}\otimes\mathscr{X})$$

为一个 $d_1\times d_2\times\cdots\times d_p\times d_1\times\cdots\times d_p$ 数组。那么对于任意 $\mathscr{W}\in\Theta$,

$$\mathrm{var}(\langle\mathscr{X},\mathscr{W}\rangle)=\langle\Sigma,\mathscr{W}\otimes\mathscr{W}\rangle=\langle\Sigma,\mathbf{w}^{(1)}\otimes\cdots\otimes\mathbf{w}^{(p)}\otimes\mathbf{w}^{(1)}\otimes\cdots\otimes\mathbf{w}^{(p)}\rangle.$$

记作

$$\lambda_k=\mathrm{var}(\langle\mathscr{X},\mathscr{U}_k\rangle).$$

由于 Σ 的对称性,

$$\mathscr{U}_1\otimes\mathscr{U}_1=\mathrm{argmax}_{\mathbf{w}^{(1)}\otimes\cdots\otimes\mathbf{w}^{(2p)}:\|\mathbf{w}^{(q)}\|=1}\langle\Sigma,\mathbf{w}^{(1)}\otimes\cdots\otimes\mathbf{w}^{(2p)}\rangle$$

因此 $\lambda_1 \mathscr{U}_1 \otimes \mathscr{U}_1$ 也是 Σ 的最佳秩一近似（例如，参见 Friedland，2013）。类似地，

$$\mathscr{U}_k \otimes \mathscr{U}_k = \operatorname{argmax}_{\substack{\mathbf{w}^{(1)} \otimes \cdots \otimes \mathbf{w}^{(2p)} : \|\mathbf{w}^{(q)}\| = 1 \\ \mathbf{w}^{(1)} \otimes \cdots \otimes \mathbf{w}^{(2p)} \perp_c \mathscr{U}_l \otimes \mathscr{U}_l, \quad l<k}} \langle \Sigma, \mathbf{w}^{(1)} \otimes \cdots \otimes \mathbf{w}^{(2p)} \rangle$$

在向量情形中，例如 $p = 1$，$\{(\lambda_k, \mathscr{U}_k) : k \geq 1\}$ 是协方差矩阵Σ 的特征对，并且

$$\Sigma_r := \sum_{k=1}^{r} \lambda_k \mathscr{U}_k \otimes \mathscr{U}_k$$

是对 Σ的最佳秩-r 近似，即

$$\Sigma_r = \operatorname{argmin}_{A \in \mathbb{R}^{d_1 \times d_1} : \operatorname{rank}(A) \leq r} \|A - \Sigma\|.$$

当 $p > 1$时，这种刻画就变得不可靠，因为最佳低秩近似的概念本身也变得不确定。对于矩阵，得益于 Eckart-Young定理，最佳低秩近似可以通过奇异值分解确定。张量的低秩近似要微妙得多，一般而言，最佳低秩近似可能并不存在。例如，参见 Hackbusch(2012)。尽管如此，根据构造，Σ_r 是对r 的所谓最佳秩贪心正交近似，Σ。例如，参见Kolda(2001)。特别地，当多路结构确实以一种使通常的主成分为秩一张量的方式显现时，例如，(1)和(4)的解为秩一，此时 Σ_r 是对 Σ的最佳低秩近似。

样本多向主成分也可以用类似的方式定义。具体而言，给定由 $\mathscr{X}$ 的独立副本组成的样本 $\mathscr{X}_1, \ldots, \mathscr{X}_n$，可以通过最大化样本方差来估计 $\mathscr{U}_k$s:

$$\widehat{\mathscr{U}_k} := \operatorname{argmax}_{\mathscr{W} \in \Theta : \mathscr{W} \perp_c \mathscr{U}_l, \forall l<k} \frac{1}{n} \sum_{i=1}^{n} \langle \mathscr{X}_i, \mathscr{W} \rangle^2. \tag{6}$$

Let

$$\widehat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \mathscr{X}_i \otimes \mathscr{X}_i$$

$\widehat{\ }$为样本协方差算子。那么，$\mathscr{U}_1$ 可以通过对 Σ 进行最佳秩一近似来定义-$\widehat{\ }$ mation

$$\widehat{\mathscr{U}_1} = \operatorname{argmax}_{\mathscr{W} \in \Theta} \langle \widehat{\Sigma}, \mathscr{W} \otimes \mathscr{W} \rangle.$$

其他主成分也可以等价地定义为

$$\widehat{\mathscr{U}_k} = \operatorname{argmax}_{\mathscr{W} \in \Theta : \mathscr{W} \perp_c \widehat{\mathscr{U}_l}, \forall l<k} \langle \widehat{\Sigma}, \mathscr{W} \otimes \mathscr{W} \rangle.$$

$\widehat{\ }$注意，$\mathscr{U}_k$ 也可以被视为经过消去处理的协方差算子的最佳秩一近似:

$$\widehat{\mathscr{U}_k} = \text{argmax}_{\mathscr{W}\in\Theta}\langle\check{\Sigma}_k, \mathscr{W}\otimes\mathscr{W}\rangle,$$

其中

$$\check{\Sigma}_k = \widehat{\Sigma}\times_1\widehat{\mathcal{P}}_k^{(1)}\times_2\cdots\times_p\widehat{\mathcal{P}}_k^{(p)}\times_{p+1}\widehat{\mathcal{P}}_k^{(1)}\times_{p+2}\cdots\times_{2p}\widehat{\mathcal{P}}_k^{(p)}$$

且 $\widehat{\ }\mathcal{P}_k^{(1)}$ 是由$\{\widehat{\ }\mathbf{u}_l^{(q)}: 1\le l<k\}$张成的线性子空间的投影矩阵。下文中，$\times_q$ 表示张量 $\mathscr{T}\in\mathbb{R}^{d_1\times d_2\times\ldots\times d_k}$ 与矩阵 $A\in\mathbb{R}^{m\times d_q}$ 之间的模态 q 乘积，因此 $\mathscr{T}\times_q A\in\mathbb{R}^{d_1\times\ldots d_{q-1}\times m\times d_{q+1}\ldots\times d_k}$ 的元素为

$$[\mathscr{T}\times_q A]_{i_1\ldots i_{q-1}ji_{q+1}\ldots i_k} = \sum_{i_q=1}^{d_q}\mathscr{T}_{i_1\ldots i_q\ldots i_k}A_{ji_q}.$$

求张量的最佳秩一近似是数值线性代数中的经典问题，而将样本多路主成分分析表述为这一问题，使我们能够利用现有的众多算法来求解。本文关注多路主成分分析的统计性质。对计算方面的进一步讨论感兴趣的读者可参阅，例如，Zhang 和 Golub (2001); Hackbusch(2012); Janzamin 等人(2019)及其中的参考文献。

与通常的主成分类似，多路主成分可用于构造原始数据的低秩近似。不过，这两类主成分分析之间也存在根本性的差异，尽管有时这些差异并不明显。经过适当中心化后，通常的样本主成分与数据矩阵的主奇异向量一致，因此可以通过奇异值分解（SVD）计算。相比之下，尽管多路主成分分析与原始数据数组的最佳低秩近似密切相关，但一般而言，两者并不等同。具体来说，考虑将观测堆叠成一个高阶张量$\mathbf{X}\in\mathbb{R}^{n\times d_1\times\ldots\times d_p}$，其第 i 个正面切片为 $\mathscr{X}_i$。当 $\mathscr{X}$ 是一个向量时，即 $p=1$,$\widehat{\ }\mathbf{X}$ 是一个矩阵，而如上所定义的样本主成分 $\mathscr{U}_k$ 是其第 k 个右奇异向量。因此，很自然地会想到采用相同的方法，直接通过求 $\mathbf{X}$ 的最佳正交低秩近似来估计 $\mathscr{U}_k$:

$$\min_{\substack{\mathscr{W}_1,\ldots,\mathscr{W}_r\in\Theta,\mathbf{a}_1,\ldots,\mathbf{a}_r\in\mathbb{R}^n\\ \mathscr{W}_l\perp_c\mathscr{W}_k,\mathbf{a}_l\perp\mathbf{a}_k,\forall l\neq k}}\|\mathbf{X}-(\mathbf{a}_1\otimes\mathscr{W}_1+\cdots+\mathbf{a}_r\otimes\mathscr{W}_r)\|_{\text{F}} \tag{7}$$

例如，参见 Harshman 和 Lundy (1984)。这个问题通常称为张量奇异值分解问题，近年来受到了广泛关注。例如，参见Richard 和 Montanari (2014)；

Hopkins 等（2015）；Liu 等（2017）；Zhang 和 Xia（2018）；Auddy 和 Yuan（2020）。然而，样本多路主成分通常并不是（7）的解。首先，X 的最佳正交秩-r 近似与秩-$(r-1)$近似之间的差通常不是秩一张量，因此无法与多路主成分对应。例如，参见 Hackbusch（2012）。为克服这一难题，可以考虑以贪心方式求解（7），也就是说，在固定其他向量的情况下，仅对 $\mathscr{W}_k$ 和 $\mathbf{a}_k$ 进行优化。然而，一般而言，这仍会得到一组不同的主成分，因为（7）对 $\mathbf{a}_k$s 施加了额外的正交约束。正如我们将看到的，多路主成分分析与数据张量低秩近似之间的这种细微区别，不仅意味着多路主成分分析需要采用不同于张量奇异值分解的处理方法，也会导致两者呈现不同的统计特性。

3 收敛速率

一个自然的问题是，$\mathscr{U}_k$ 及其分量 $\mathbf{u}_k^{(q)}$ s 能否通过样本对应量得到良好估计。现在我们将探讨这个问题，并研究样本多路主成分的收敛速率。一方面，除了可解释性更强之外，我们还将进一步论证多路主成分分析优于将数据串联后进行传统主成分分析。另一方面，我们的研究也带来了对样本多路主成分分析运行特性的新认识，并揭示了它与传统主成分分析之间耐人寻味的差异。为明确起见，我们将考虑所谓的尖峰协方差模型，作为理论推导的工作模型。

假设随机数组 $\mathscr{X} \in \mathbb{R}^{d_1 \times \ldots \times d_p}$ 服从线性因子模型：

$$\mathscr{X} = \sum_{k=1}^{r} \sigma_k \theta_k \mathscr{U}_k + \sigma_0 \mathscr{E}, \tag{8}$$

其中 $(\theta_1, \ldots, \theta_r)^\top \sim N(\mathbf{0}, I_r)$ 为随机因子载荷，$\mathscr{U}_k$s $(\in \Theta)$ 为单位长度的秩一主成分，满足 $\mathscr{U}_k \perp_c \mathscr{U}_l$ 对任意 $k \neq l$均成立；$\mathscr{E}$ 为一个各元素相互独立且服从 $N(0,1)$ 分布的噪声张量。值得指出的是，我们的结果和论证可以推广到正态分布之外，并适用于一般的次高斯分布。为便于表述，我们采用正态性假设。不失一般性，我们还假设信号的特征值均非零，并按非递增顺序排列，

即 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$。注意，我们不要求 σ_ks 互不相同。不难看出，上述 $\mathscr{X}$ 的协方差算子为

$$\Sigma = \sum_{k=1}^{r} \sigma_k^2 \mathscr{U}_k \otimes \mathscr{U}_k + \sigma_0^2 \mathscr{I},$$

其中 $\mathscr{I}$ 是单位张量，即 $\mathscr{I}_{j_1 \ldots j_p j'_1 \ldots j'_p} = 1$ 若 $j_q = j'_q$ 对所有 $q = 1, \ldots, p$ 则取该值，否则为 0。尖峰协方差模型（例如（8））被广泛用作工作模型，用于研究向量观测情况下的主成分分析，即 $p = 1$。例如，参见 Johnstone (2001) 和Paul (2007)。

本节将建立样本多路主成分的收敛速率。为此，记 $\angle(\mathbf{w}_1,\mathbf{w}_2)$ 为两个向量 $\mathbf{w}_1$ 和 $\mathbf{w}_2$ 之间的夹角，其取值范围为$[0, \pi/2]$；类似地，对于两个数组 $\mathscr{W}_1$ 和 $\mathscr{W}_2$，$\angle(\mathscr{W}_1, \mathscr{W}_2)$ 表示它们的向量化 $\mathrm{vec}(\mathscr{W}_1)$ 和 $\mathrm{vec}(\mathscr{W}_2)$ 之间的夹角。

先从经典情形入手颇有启发性：维数 $d_1, \ldots, d_p$ 以及所有其他参数（例如 σ_0、σ_ks 和 r）均保持固定，而样本量 n 趋于无穷。我们的第一个结果表明，在此情形下，样本主成分 $\widehat{\mathscr{U}_k}$ 及其分量 $\hat{}\mathbf{u}_k^{(q)}$ s 具有根-n 一致性。

定理 3.1. 设 $\mathscr{X}_1, \ldots, \mathscr{X}_n$ 是服从尖峰协方差模型 *(8)* 的独立观测，其中 $p > 1$ 满足 $\mathscr{U}_k = \boldsymbol{u}_k^{(1)} \otimes \cdots \otimes \boldsymbol{u}_k^{(p)}$，且 $\sigma_k > 0$。假设所有参数均固定，样本量 n 不断增加。设 $\widehat{\mathscr{U}_k} = \hat{}\boldsymbol{u}_k^{(1)} \otimes \cdots \otimes \hat{}\boldsymbol{u}_k^{(p)}$ 为按 *(6)* 定义的样本多路主成分。则存在一个作用于 $[r] := \{1, \ldots, r\}$ 上的置换 π，使得

$$\max_{1 \leq q \leq p} \sin \angle(\hat{\mathbf{u}}_k^{(q)}, \mathbf{u}_{\pi(k)}^{(q)}) = O_p(n^{-1/2}), \tag{9}$$

对所有 $k \in [r]$ 均成立，因此

$$\sin \angle(\widehat{\mathscr{U}_k}, \mathscr{U}_{\pi(k)}) = O_p(n^{-1/2}), \qquad k = 1, \ldots, r$$

当 $n \to \infty$ 时。

上述结果与传统主成分分析（例如，Anderson，1984）的结果之间最显著的区别在于，样本多向主成分的根-n 一致性并不要求协方差矩阵的特征值（$(\sigma_k^2 + \sigma_0^2)$ 或等价地 σ_k）是单重的，即 $\sigma_k \neq \sigma_{k+1}$。请注意，如果没有多向结构约束，传统

主成分只有在其对应的特征值为单重时才有唯一确定的定义，因此才可能被估计。正如定理 3.1所示，多路主成分分析不需要这一限制。对于多路主成分分析，无论其他特征值如何，每个样本主成分都是根-n 一致的。还值得注意的是，由于我们不要求 σ_k彼此不同，因此无法保证 $\widehat{\mathscr{U}}_k$ 估计量 $\mathscr{U}_k$。这并非多路主成分分析的缺陷，而是由于第 k大特征值可能存在不确定性所致。事实上，如果 $\sigma_{k+1} < \sigma_k < \sigma_{k-1}$，那么我们可以在定理 $\pi(k) = k$ 中选择 3.1。一般而言，定理 3.1表明，每个样本主成分必然是某个多路主成分的根-n 一致估计。

为了进一步理解多路主成分分析的运行特性和优点，我们现在考察更一般的情形，并进一步凸显维数和信噪比的作用。为简洁起见，以下假设 $\mathscr{X}$ 是“近似立方”的，即存在常数 $0 < c_1, \ c_2 < \infty$ 使得 $c_1 d \le d_1, \ldots, d_p \le c_2 d$ 对某个自然数 d 成立，且该数可能随 n趋于无穷。一般情形也可类似处理，但记号会复杂得多，推导也更加繁琐。

定理 3.2。设 $\mathscr{X}_1, \ldots, \mathscr{X}_n$ 是服从尖峰协方差模型 (8) 的独立观测，其中 $p > 1$ 满足 $\mathscr{U}_k = \mathbf{u}_k^{(1)} \otimes \cdots \otimes \mathbf{u}_k^{(p)}$。设 $\widehat{\mathscr{U}}_k = \widehat{\mathbf{u}}_k^{(1)} \otimes \cdots \otimes \widehat{\mathbf{u}}_k^{(p)}$ 为由 (6)定义的样本多路主成分。假设

$$r \log r \le c_0 \min\{n, d\} \quad \text{and} \quad \left(\frac{\sigma_0}{\sigma_r} + \frac{\sigma_0^2}{\sigma_r^2}\right) \cdot \max\left\{\sqrt{\frac{d}{n}}, \frac{d}{n}\right\} \le \frac{c_0}{\sqrt{r}}, \tag{10}$$

对于足够小的常数 $c_0 > 0$，存在常数 $C > 0$ 和关于 $[r]$ 的置换 π，使得

$$\max_{1 \le q \le p} \sin \angle(\widehat{\mathbf{u}}_k^{(q)}, \mathbf{u}_{\pi(k)}^{(q)}) \le C \left(\frac{\sigma_0}{\sigma_{\pi(k)}} + \frac{\sigma_0^2}{\sigma_{\pi(k)}^2}\right) \cdot \max\left\{\sqrt{\frac{d}{n}}, \frac{d}{n}\right\}, \tag{11}$$

对所有 $k \in [r]$ 均成立，因此

$$\sin \angle(\widehat{\mathscr{U}}_k, \mathscr{U}_{\pi(k)}) \le C \left(\frac{\sigma_0}{\sigma_{\pi(k)}} + \frac{\sigma_0^2}{\sigma_{\pi(k)}^2}\right) \cdot \max\left\{\sqrt{\frac{d}{n}}, \frac{d}{n}\right\}, \qquad k = 1, \ldots, r,$$

当 n 趋于无穷时，该结论以趋于 *1* 的概率成立。

定理3.2 可以视为定理3.1的推广。其证明相当复杂，我们将简要讨论其中的一些挑战，以及解决这些挑战的主要思路。

证明通过对 k 进行归纳法展开。若特征值不是单重的，或特征值间隙较小，则需要特别处理。这使得难以确定样本多路主成分估计的是哪个多路主成分，或者等价地，置换 π。为此，我们将定义

$$\pi(1) = \mathrm{argmax}_{1\le l\le r}\left\{\sigma_l^2\left|\prod_{q=1}^{p}\langle \mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_1^{(q)}\rangle\right|\right\},$$

以及对于 $k>1$,

$$\pi(k) := \mathrm{argmax}_{l\notin\pi([k-1])}\left\{\sigma_l^2\left|\prod_{q=1}^{p}\langle \mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_k^{(q)}\rangle\right|\right\}.$$

为了彻底消除特征值间隙的影响，我们需要仔细量化 $\widehat{\mathscr{U}}_1,\ldots,\widehat{\mathscr{U}}_{k-1}$ 的估计误差对第 k 个样本多路主成分的影响。为此，我们将分别推导二者的界

$$\max_{1\le q\le p}\sin\angle(\mathbf{u}_{\pi(k)}^{(q)}, \widehat{\mathbf{u}}_k^{(q)}),$$

and

$$\max_{1\le q\le p}\max_{l\notin\pi([k])}\langle \mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_k^{(q)}\rangle,$$

并利用后者可能远小于前者这一事实。

当 $d=O(n)$ 时，定理 3.2 给出的收敛速率为

$$\sin\angle(\widehat{\mathscr{U}}_k, \mathscr{U}_{\pi(k)}) \le C\left(\frac{\sigma_0}{\sigma_{\pi(k)}}+\frac{\sigma_0^2}{\sigma_{\pi(k)}^2}\right)\cdot\sqrt{\frac{d}{n}};$$

当 $d\gg n$ 时，我们有

$$\sin\angle(\widehat{\mathscr{U}}_k, \mathscr{U}_{\pi(k)}) \le C\left(\frac{\sigma_0}{\sigma_{\pi(k)}}+\frac{\sigma_0^2}{\sigma_{\pi(k)}^2}\right)\cdot\frac{d}{n}.$$

特别地，$\widehat{\mathscr{U}}_k$ 是一致的，例如，

$$\sin\angle(\widehat{\mathscr{U}}_k, \mathscr{U}_{\pi(k)}) = o_p(1),$$

只要 $\sigma_{\pi(k)}/\sigma_0 \gg \max\{d/n, (d/n)^{1/4}\}$。

这里特别值得关注的是维数的作用。定理 3.2给出的收敛速率取决于维数 d，而非环境维度$D := d_1d_2\cdots d_p$。这是因为多路主成分分析将主成分限制为秩一张量，因此参数更少。这种降维对于

多路数据尤为重要。例如，考虑 σ_0 和 σ_k固定的情况，根据Johnstone和Lu(2009）的结果，直接将数据串联后应用通常的主成分分析，必然会导致对 $\mathscr{U}_k$ 的不一致估计，只要 $D \gg n$。然而，我们的结果表明，只要 $d \ll n$，多路主成分分析就是一致的。

为了进一步与传统主成分分析进行比较，我们现在关注以下情形： $d \ll n$以及 r、 σ_0、σ_1、……、 σ_r 均为固定值。如Birnbaum 等人（2013）所示，在这一条件下，传统主成分分析（即 $p = 1$）满足

$$\sin \angle(\widehat{\mathscr{U}_k}, \mathscr{U}_{\pi(k)}) \asymp \left(\frac{\sigma_0}{\sigma_{\pi(k)}} + \frac{\sigma_0^2}{\sigma_{\pi(k)}^2} \right) \cdot \sqrt{\frac{d}{n}} + \frac{1}{\sqrt{n}} \left(\sum_{k' \neq k} \frac{(\sigma_0 + \sigma_{\pi(k)})(\sigma_0 + \sigma_{\pi(k')})}{\sigma_{\pi(k)}^2 - \sigma_{\pi(k')}^2} \right)$$

以趋于 1 的概率成立。将上述速率与定理 3.2中的速率进行比较，可以清楚地看出，两者的差异在于右端的第二项。传统主成分分析中存在这一项，意味着 σ_k之间不应存在并列。即使σ_k彼此均不相同，我们对一个主成分的估计效果也在很大程度上取决于其对应特征值与其他特征值之间的间隙，当 $p = 1$时尤其如此。相比之下，定理 3.2 给出的界仅由$\sigma_{\pi(k)}$决定，而不像传统主成分分析的情形那样取决于特征值间隙 $\min\{\sigma_{\pi(k)-1}^2 - \sigma_{\pi(k)}^2, \sigma_{\pi(k)}^2 - \sigma_{\pi(k)+1}^2\}$。

将定理3.2 中的多路主成分分析收敛速率与张量奇异值分解的收敛速率进行比较也很有启发性。回顾一下，

$$\mathbf{X} = \sum_{k=1}^{r} \sigma_k \Theta_k \otimes \mathscr{U}_k + \sigma_0 \mathbf{E},$$

其中 $\Theta_k = (\theta_{1k}, \ldots, \theta_{nk})^\top$ 是一个包含 θ_k 的 n 个实现值的向量，而 $\mathbf{E}$ 是一个 $n \times d_1 \times \cdots \times d_p$ 张量，其第 i 个正面切片为 $\mathscr{E}_i$。相比之下，在张量 SVD 模型中， Θ_k 是确定性的。如果 Θ_k 彼此正交，那么 $\mathscr{U}_k$ 可以以 σ_0kEk/(σ_kkΘ_kk) 的速率进行估计，其阶为 (σ_0/σ_k) max{$\sqrt{d/n}$, 1}。这直接源于 Auddy 和 Yuan（2020）的扰动界；Richard and Montanari (2014) 也在秩一情形下，即 $r = 1$，推导出了类似的界。然而，在我们的情形中， Θ_k 和 Θ_l 是随机的，通常并不彼此正交。因此，我们得到的速率在其对信噪比 σ_k/σ_0 的依赖性方面有所不同。对于传统主成分分析，也观察到了类似现象（例如，参见 Birnbaum 等，2013 年）。

4 渐近正态性与偏差校正

现在我们转向多路主成分分析的分布性质。这要求我们进一步阐明偏差在样本主成分中的作用。众所周知，当维度（D）相对于样本量较大时，传统主成分分析存在偏差。例如，参见 Koltchinskii 和 Lounici（2014）；Koltchinskii 等（2020）及其中的参考文献。对于样本多路主成分，也存在同样的现象：当每个模态的维度（d）相对于样本量较大时，会产生不可忽略的偏差。此外，由于在缺乏特征值间隙时多路主成分的排序存在歧义，样本多路主成分还会产生一个更为微妙的偏差来源。如前所述，缺乏特征值间隙意味着第 k 个主成分不一定由第 k 个样本多路主成分估计。作为一个更具体的例子，考虑 $r=2$ 和 $\lambda_1=\lambda_2$ 的情形。此时，$\mathscr{U}_1$ 可以分别由 $\mathscr{U}_1$ 或 $\mathscr{U}_2$ 进行估计，而正如定理 3.2 所示，在这两种情形下，收敛速率 $\widehat{\ }$ 保持不变。但是，两种情形的渐近分布可能不同：$\mathscr{U}_2$ 必须与 $\mathscr{U}_1$ 正交，而用 $\mathscr{U}_2$ 估计 $\mathscr{U}_1$ 可能会引入额外偏差。

本节将介绍校正两种偏差的方法，并建立偏差校正后的样本主成分的渐近正态性。按照文献中的惯例，为简洁起见，我们假设 r 和 σ_1、……、σ_r 均为固定值。根据上一节的结果，在此设定下，若 $d \ll n$，则样本主成分是一致的。因此，本节将重点讨论这一情形。

4.1 当 $d=o$（$\sqrt{n}$） $\bar{\ }$

当 d 不太大时，偏差完全源于特征值可能重复，从而导致主成分的排序存在歧义。事实上，如果 σ_k互不相同，那么当 $d=o(\sqrt{n})$ 时就无需进行偏差校正，而且本小节的所有结果都适用于样本多路主成分。但在实践中，我们可能不知道或不愿假设特征值都是单重的。幸运的是，只需对样本主成分进行简单的单步更新，就能轻松消除任何可能的偏差。更具体地说，我们将考虑估计$\mathbf{u}_{\pi\ k}^{(q)}()$，即$\tilde{\ }\mathbf{u}_k^{(q)}$ 的主特征向量，其中

$$\widehat{\Sigma}(\widehat{\mathbf{u}}_k^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(q-1)},\cdot,\widehat{\mathbf{u}}_k^{(q+1)},\ldots,\widehat{\mathbf{u}}_k^{(p)},\widehat{\mathbf{u}}_k^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(q-1)},\cdot,\widehat{\mathbf{u}}_k^{(q+1)},\ldots,\widehat{\mathbf{u}}_k^{(p)}).$$

额外的这一步解除了施加于第 k 个样本多路主成分的正交约束，因此可以抑制 $\mathscr{U}_1, \ldots,$ $\mathscr{U}_{k-1}$ 的任何不利影响。

现在考虑偏差校正后的样本主成分的渐近分布。我们再次从经典情形开始，即当所有参数固定且 n 增大时。

定理 4.1。 设 $\mathscr{X}_1$，……， $\mathscr{X}_n$ 是服从尖峰协方差模型 （8）的独立观测，其中 $p>1$ 满足 $\mathscr{U}_k=\mathbf{u}_k^{(1)} \otimes \cdots \otimes \mathbf{u}_k^{(p)}$，并且 $\sigma_k>0$。假设当样本量 n 增大时，所有参数均固定。设 $\tilde{}$ $\mathbf{u}_1^{(q)}$，……， $\tilde{}\mathbf{u}_r^{(q)}$如上定义。则存在一个置换 π： $[r] \to [r]$ 使得

$$\sqrt{n}\left[\operatorname{vec}(\tilde{\mathbf{U}}^{(q)})-\operatorname{vec}(\mathbf{U}_\pi^{(q)})\right]$$
$$\xrightarrow{d} N\left(0, \operatorname{diag}\left(\left(\frac{\sigma_0^2}{\sigma_{\pi(1)}^2}+\frac{\sigma_0^4}{\sigma_{\pi(1)}^4}\right)\mathcal{P}_{\mathbf{u}_{\pi(1)}^{(q)}}^{\perp}, \ldots,\left(\frac{\sigma_0^2}{\sigma_{\pi(r)}^2}+\frac{\sigma_0^4}{\sigma_{\pi(r)}^4}\right)\mathcal{P}_{\mathbf{u}_{\pi(r)}^{(q)}}^{\perp}\right)\right),$$

当 $n \to \infty$时，其中 $\tilde{}\mathbf{U}^{(q)}=[\tilde{}\mathbf{u}_1^{(q)}$、……、 $\tilde{}\mathbf{u}_r^{(q)}]$、$\mathbf{U}_\pi^{(q)}=[\mathbf{u}_{\pi\ 1}^{(q)}(\ )\ \cdots$、$\mathbf{u}_{\pi\ r}^{(q)}(\ )]$ 以及 $\boldsymbol{P}^{\perp}{}_{\mathbf{u}_k^{(q)}}=I_{d_q}-\mathbf{u}_k^{(q)} \otimes \mathbf{u}_k^{(q)}$。

定理4.1 表明

$$n \cdot \operatorname{var}\left(\tilde{\mathbf{u}}_k^{(q)}\right) \to \frac{\sigma_0^2(\sigma_{\pi(k)}^2+\sigma_0^2)}{\sigma_{\pi(k)}^4}\left(I-\mathbf{u}_{\pi(k)}^{(q)} \otimes \mathbf{u}_{\pi(k)}^{(q)}\right),$$

and

$$n \cdot \operatorname{cov}\left(\tilde{\mathbf{u}}_k^{(q)}, \tilde{\mathbf{u}}_l^{(q)}\right) \to 0.$$

也就是说，所有多路主成分的估计量都渐近正态，且彼此独立。还要注意， $\tilde{}\mathbf{u}_k^{(q)}$ 的渐近分布不依赖于其他特征值或主成分。换言之，即使已知所有其他分量 $\mathscr{U}_l, l \neq k$ ，其估计精度也相同！这与传统主成分分析形成对照，在传统主成分分析中，$\mathbf{u}_k^{(q)}$ 的渐近分布取决于所有其他特征向量和特征值。

更具体地说，众所周知，在向量情形下，即当 $p=1$时，在附加假设$\sigma_1^2, \ldots, \sigma_r^2$互不相同的条件下，样本主成分满足

$$n \cdot \operatorname{var}\left(\hat{\mathbf{u}}_k^{(1)}\right)$$
$$\to \sum_{1 \leq l \leq r, l \neq k} \frac{(\sigma_k^2+\sigma_0^2)(\sigma_l^2+\sigma_0^2)}{(\sigma_k^2-\sigma_l^2)^2} \mathbf{u}_l^{(1)} \otimes \mathbf{u}_l^{(1)}+\frac{(\sigma_k^2+\sigma_0^2)\sigma_0^2}{\sigma_k^4}\left(I-\sum_{1 \leq l \leq r} \mathbf{u}_l^{(1)} \otimes \mathbf{u}_l^{(1)}\right)$$

且对于任意 $1 \leq l \leq r$ 和 $l \neq k$,

$$n \cdot \operatorname{cov}\left(\hat{\mathbf{u}}_k^{(1)}, \hat{\mathbf{u}}_l^{(1)}\right) \to -\frac{(\sigma_k^2+\sigma_0^2)(\sigma_l^2+\sigma_0^2)}{(\sigma_k^2-\sigma_l^2)^2} \cdot \mathbf{u}_k^{(1)} \otimes \mathbf{u}_l^{(1)}.$$

例如，参见Anderson (1984)。显然，样本主成分总是彼此相关。此外，注意到

$$\frac{\sigma_0^2(\sigma_k^2+\sigma_0^2)}{\sigma_k^4} \leq \frac{(\sigma_k^2+\sigma_0^2)(\sigma_l^2+\sigma_0^2)}{(\sigma_k^2-\sigma_l^2)^2}.$$

对于任意 $k \neq l \leq r$，严格不等式均成立。这表明，与具有相同特征值的常规主成分相比，估计得到的多路主成分变异更小。

现在我们将注意力转向更一般的情形，其中维数和其他参数允许随 n 发散。由于不同样本量下的主成分维数可能不同，因此考虑其线性形式更为自然，例如 $\langle \mathbf{u}_k^{(q)}, \mathbf{v}\rangle$，其中 $\mathbf{v} \in \mathbb{R}^{d_q}$ 是某个固定向量。若维数固定，则定理 4.1 立即表明，$\langle \tilde{\mathbf{u}}_k^{(q)}, \mathbf{v}\rangle$ 估计了 $\langle \mathbf{u}_{\pi(k)}^{(q)}, \mathbf{v}\rangle$，并且

$$\sqrt{n}\left(\langle \tilde{\mathbf{u}}_k^{(q)}, \mathbf{v}\rangle - \langle \mathbf{u}_{\pi(k)}^{(q)}, \mathbf{v}\rangle\right) \to_d N\left(0, \left(\frac{\sigma_0^2}{\sigma_{\pi(k)}^2} + \frac{\sigma_0^4}{\sigma_{\pi(k)}^4}\right) \|\mathcal{P}_{\mathbf{u}_{\pi(k)}^{(q)}}^{\perp} \mathbf{v}\|^2\right)$$

以下结果表明，只要 $d = o(\sqrt{n})$，这一结论仍然成立。

定理 4.2。 设 $\mathscr{X}_1, \ldots\ldots, \mathscr{X}_n$ 是遵循尖峰协方差模型 (8) 的独立观测，且 $p > 1$ 满足 $\mathscr{U}_k = \mathbf{u}_k^{(1)} \otimes \cdots \otimes \mathbf{u}_k^{(p)}$，且 $\sigma_k > 0$。假设 (10) 成立，$d = o(\sqrt{n})$。那么存在一个置换 $\pi : [r] \to [r]$ 使得

$$\sqrt{n}\left(\langle \tilde{\mathbf{u}}_k^{(q)}, \mathbf{v}\rangle - \langle \mathbf{u}_{\pi(k)}^{(q)}, \mathbf{v}\rangle\right) \to_d N\left(0, \left(\frac{\sigma_0^2}{\sigma_{\pi(k)}^2} + \frac{\sigma_0^4}{\sigma_{\pi(k)}^4}\right) \|\mathcal{P}_{\mathbf{u}_{\pi(k)}^{(q)}}^{\perp} \mathbf{v}\|^2\right)$$

如 $n \to \infty$。

定理 4.2 表明，只要满足 $\tilde{\mathbf{u}}_j^{(q)}$（$d = o(\sqrt{n})$），就可以预期其渐近行为与固定维数情形相同。

4.2 当 $d = o(n)$

对于更高维情形，上文所述的简单偏差校正已不再适用。仔细考察可知，当 $\tilde{\mathbf{u}}_j^{(q)}$ 时，仍会产生不可忽略的偏差。值得庆幸的是，受 Koltchinskii 和 Lounici(2014) 针对传统主成分分析提出的方案启发，我们可以采用类似的样本拆分方法校正这两类偏差。

不失一般性，假设 n 是偶数，并将n 个观测随机分成两半：$(\mathscr{X}_1, \ldots, \mathscr{X}_{n/2})$ 和 $(\mathscr{X}_{n/2+1}, \ldots, \mathscr{X}_n)$。分别记为 $\hat{\Sigma}^{[1]}$ 和

$\widehat{\Sigma}^{[2]}$ 基于两半数据得到的样本协方差算子。同样，我们将记 $\widehat{\mathscr{U}_k^{[h]}} = \widehat{\mathbf{u}_k^{(1)\,[h]}} \otimes \cdots \otimes \widehat{\mathbf{u}_k^{(p)\,[h]}}$，基于数据的 h （$= 1$ 或 2）半得到的第 k个样本主成分。然而，如前所述，$\widehat{\mathscr{U}_k^{[1]}}$ 和 $\widehat{\mathscr{U}_k^{[2]}}$ 可能并不估计同一个主成分。为此，我们将以 $\widehat{\mathscr{U}_k}$（即基于完整数据集得出的估计量）为参照，对 $\widehat{\mathscr{U}_k^{[1]}}$ 和 $\widehat{\mathscr{U}_k^{[2]}}$ 重新排序。具体而言，不失一般性，我们假设

$$\widehat{\mathscr{U}_k^{[1]}} = \operatorname{argmin}_{k \le l \le r} \sin \angle \left(\widehat{\mathscr{U}_l^{[1]}}, \widehat{\mathscr{U}_k} \right).$$

对 $\widehat{\mathscr{U}_k^{[2]}}$ s 采用相同的重新标记步骤。另请注意，主成分的符号无关紧要，因为 $\mathscr{U}_k$ 和 $-\mathscr{U}_k$ 表示相同的变换。因此，以下也不失一般性地假设$\langle \widehat{\mathbf{u}_k^{(q)\,[1]}}, \widehat{\mathbf{u}_k^{(q)\,[2]}} \rangle \ge 0$。

回顾一下，

$$\sigma_k^2 \mathbf{u}_k^{(q)} \mathbf{u}_k^{(q)\top} + \sigma_0^2 I_{d_q} = \Sigma(\mathbf{u}_k^{(1)}, \ldots, \mathbf{u}_k^{(q-1)}, \cdot, \mathbf{u}_k^{(q+1)}, \ldots, \mathbf{u}_k^{(p)}, \mathbf{u}_k^{(1)}, \ldots, \mathbf{u}_k^{(q-1)}, \cdot, \mathbf{u}_k^{(q+1)}, \ldots, \mathbf{u}_k^{(p)}).$$

然后，我们将利用上述恒等式更新样本主成分，其中 Σ 和$\mathbf{u}_k^{(q)}$ s 是分别从两半样本中估计得到的。记 $\check{\mathbf{u}}_k^{(q)\,[1]}$，为以下矩阵的主特征向量：

$$\widehat{\Sigma}^{[1]}(\widehat{\mathbf{u}}_k^{(1),[2]}, \ldots, \widehat{\mathbf{u}}_k^{(q-1),[2]}, \cdot, \widehat{\mathbf{u}}_k^{(q+1),[2]}, \ldots, \widehat{\mathbf{u}}_k^{(p),[2]}, \widehat{\mathbf{u}}_k^{(1),[2]}, \ldots, \widehat{\mathbf{u}}_k^{(q-1),[2]}, \cdot, \widehat{\mathbf{u}}_k^{(q+1),[2]}, \ldots, \widehat{\mathbf{u}}_k^{(p),[2]}),$$

类似地，$\check{\mathbf{u}}_k^{(q)\,[2]}$，为以下矩阵的主特征向量：

$$\widehat{\Sigma}^{[2]}(\widehat{\mathbf{u}}_k^{(1),[1]}, \ldots, \widehat{\mathbf{u}}_k^{(q-1),[1]}, \cdot, \widehat{\mathbf{u}}_k^{(q+1),[1]}, \ldots, \widehat{\mathbf{u}}_k^{(p),[1]}, \widehat{\mathbf{u}}_k^{(1),[1]}, \ldots, \widehat{\mathbf{u}}_k^{(q-1),[1]}, \cdot, \widehat{\mathbf{u}}_k^{(q+1),[1]}, \ldots, \widehat{\mathbf{u}}_k^{(p),[1]}).$$

为避免样本拆分造成效率损失，我们考虑一个新的估计量 $\check{\mathscr{U}}_k = \check{\mathbf{u}}_k^{(1)} \otimes \cdots \otimes \check{\mathbf{u}}_{(p)\,k}$其中

$$\check{\mathbf{u}}_k^{(q)} = \frac{\check{\mathbf{u}}_k^{(q),[1]} + \check{\mathbf{u}}_k^{(q),[2]}}{\left\| \check{\mathbf{u}}_k^{(q),[1]} + \check{\mathbf{u}}_k^{(q),[2]} \right\|}.$$

下面的定理表明，只要 $d = o(n^{2/3})$，我们就可以通过适当缩放$\langle \mathbf{u}_{\pi\,k\,()}^{(q)}, \mathbf{v} \rangle$ $\check{}$ $\langle \mathbf{u}_k^{(q)}, \mathbf{v} \rangle$，构造出$\langle \mathbf{u}_{\pi\,k\,()}^{(q)}, \mathbf{v} \rangle$ 的无偏估计。

定理 4.3。 设 $\mathscr{X}_1, \ldots, \mathscr{X}_n$ 是服从尖峰协方差模型 (8) 的独立观测，且 $p > 1$ 满足 $\mathscr{U}_k = \mathbf{u}_k^{(1)} \otimes \cdots \otimes \mathbf{u}_k^{(p)}$ 并且 $\sigma_k > 0$。设 $\check{}$ $\mathscr{U}_k = \check{\mathbf{u}}_k^{(1)} \otimes \cdots \otimes \check{\mathbf{u}}_k^{(p)}$

为上文定义的估计主成分。假设 r 和 $\sigma_1, \ldots, \sigma_r$ 是固定的，且 $d = o(n^{2/3})$。那么存在一个置换 $\pi : [r] \to [r]$，使得

$$\sqrt{n}\left((1+b_k^{(q)})\langle \check{\mathbf{u}}_k^{(q)}, \mathbf{v}\rangle - \langle \mathbf{u}_{\pi(k)}^{(q)}, \mathbf{v}\rangle\right) \to_d N\left(0, \left(\frac{\sigma_0^2}{\sigma_{\pi(k)}^2} + \frac{\sigma_0^4}{\sigma_{\pi(k)}^4}\right) \|\mathcal{P}^{\perp}_{\mathbf{u}_{\pi(k)}^{(q)}} \mathbf{v}\|^2\right)$$

如 $n \to \infty$ 其中

$$b_k^{(q)} = \sqrt{1 + \frac{d_q}{n}\left(\frac{\sigma_0^2}{\sigma_{\pi(k)}^2} + \frac{\sigma_0^4}{\sigma_{\pi(k)}^4}\right)} - 1. \tag{12}$$

值得指出的是，当 $d = o(n^{1/2})$时，由 (12) 描述的偏差校正因子满足 $b_k^{(q)} = o(n^{-1/2})$，因此可以忽略。这与我们先前的观察一致，也表明当 $d \ll n^{1/2}$时，样本拆分当然也是不必要的。当 $d \gg n^{1/2}$时，偏差校正就变得至关重要。特别是，定理 4.3 表明，只要 $d = o(n^{2/3})$，就可以应用显式的偏差校正因子。对于更高维的情形，类似的去偏因子是否存在显式表达式尚不清楚。尽管如此，我们仍可通过进一步的样本拆分，为所有 $d \ll n$ 推导出合适的偏差校正因子。

更具体地说，我们首先将观测随机分成两半。随后将数据的前半部分进一步分成两个大小相等的组，以计算样本协方差算子 $\widehat{\Sigma}^{[1][1]}$ 和 $\widehat{\Sigma}^{[1][2]}$，然后我们计算 $\widehat{\mathbf{u}}_k^{(q)\,[1][1]}$，和 $\widehat{\mathbf{u}}_k^{(q)\,[1][2]}$，作为以下算子的主特征向量

$$\begin{aligned}
&\widehat{\Sigma}^{[1][1]}(\widehat{\mathbf{u}}_k^{(1),[2]}, \ldots, \widehat{\mathbf{u}}_k^{(q-1),[2]}, \cdot, \widehat{\mathbf{u}}_k^{(q+1),[2]}, \ldots, \widehat{\mathbf{u}}_k^{(p),[2]},\\
&\qquad \widehat{\mathbf{u}}_k^{(1),[2]}, \ldots, \widehat{\mathbf{u}}_k^{(q-1),[2]}, \cdot, \widehat{\mathbf{u}}_k^{(q+1),[2]}, \ldots, \widehat{\mathbf{u}}_k^{(p),[2]}),\\
&\widehat{\Sigma}^{[1][2]}(\widehat{\mathbf{u}}_k^{(1),[2]}, \ldots, \widehat{\mathbf{u}}_k^{(q-1),[2]}, \cdot, \widehat{\mathbf{u}}_k^{(q+1),[2]}, \ldots, \widehat{\mathbf{u}}_k^{(p),[2]},\\
&\qquad \widehat{\mathbf{u}}_k^{(1),[2]}, \ldots, \widehat{\mathbf{u}}_k^{(q-1),[2]}, \cdot, \widehat{\mathbf{u}}_k^{(q+1),[2]}, \ldots, \widehat{\mathbf{u}}_k^{(p),[2]}),
\end{aligned}$$

同样地，我们使用数据的后半部分计算 $\widehat{\mathbf{u}}_k^{(q)\,[2][1]}$，s 和 $\widehat{\mathbf{u}}_k^{(q)\,[2][2]}$，s。与之前一样，我们将按一致的顺序和符号对这些估计值进行排序。令

$$\widehat{b}_k^{(q)} = \frac{\left\|\check{\mathbf{u}}_k^{(q),[1]} + \check{\mathbf{u}}_k^{(q),[2]}\right\|}{\sqrt{\left\langle \widehat{\mathbf{u}}_k^{(q),[1][1]}, \widehat{\mathbf{u}}_k^{(q),[1][2]}\right\rangle} + \sqrt{\left\langle \widehat{\mathbf{u}}_k^{(q),[2][1]}, \widehat{\mathbf{u}}_k^{(q),[2][2]}\right\rangle}} - 1. \tag{13}$$

定理 4.4。设 $\mathscr{X}_1$、……、$\mathscr{X}_n$ 是服从尖峰协方差模型的独立观测 (8)，且 $p > 1$ 满足 $\mathscr{U}_k = \mathbf{u}_k^{(1)} \otimes \cdots \otimes \mathbf{u}_k^{(p)}$，并且 $\sigma_k > 0$。令 $\check{}\,\mathscr{U}_k = \check{}\,\mathbf{u}_k^{(1)} \otimes \cdots \otimes \check{}\,\mathbf{u}_k^{(p)}$

为上述定义的估计主成分。假设 r 和 σ_1、……、σ_r 为固定值，且 $d = o(n)$。那么存在一个置换 π： $[r] \to [r]$，使得

$$\sqrt{n}\left((1+\widehat{b}_k^{(q)})\langle \check{\mathbf{u}}_k^{(q)}, \mathbf{v}\rangle - \langle \mathbf{u}_{\pi(k)}^{(q)}, \mathbf{v}\rangle\right) \to_d N\left(0, \left(\frac{\sigma_0^2}{\sigma_{\pi(k)}^2} + \frac{\sigma_0^4}{\sigma_{\pi(k)}^4}\right) \|\mathcal{P}^{\perp}_{\mathbf{u}_{\pi(k)}^{(q)}} \mathbf{v}\|^2\right),$$

其中 $n \to \infty$，其中 $\widehat{b}_k^{(q)}$ 由下式给出：（13）。此外，

$$\widehat{b}_k^{(q)} = \sqrt{1 + \frac{d_q}{n}\left(\frac{\sigma_0^2}{\sigma_{\pi(k)}^2} + \frac{\sigma_0^4}{\sigma_{\pi(k)}^4}\right)} - 1 + O_p\left(\frac{d^{3/2}}{n^{3/2}}\right) + o_p\left(\frac{1}{\sqrt{n}}\right).$$

根据定理 4.4，只要 $d = o$（n），就可以采用双重样本拆分方法，为多路主成分的线性形式推导置信区间。然而，这种稳健性是以增加计算成本为代价的，并且可能导致有限样本中的效率损失。在实践中，如果 d 不太大，仍可优先采用定理 4.3 所述的显式偏差校正；或者在 d 较小时采用单步更新。

4.3 关于多路主成分的推断

本节前文所展示的渐近正态性构成了对线性形式进行推断的基础：$\mathbf{h}$ $(\overset{(q)}{\pi}{}_k \mathbf{v}\rangle)$ ，。具体而言，主成分线性形式中最有趣且最简单的例子之一是其坐标，即$\mathbf{v}$ 是单位矩阵的列向量。然而，要为$\langle \mathbf{u}_{\pi}^{(q)}{}_k$（$,\mathbf{v}\rangle$）构造置信区间或检验假设，我们还需要估计其方差。具体来说，其渐近分布仅取决于 σ_0、 $\sigma_{\pi(k)}$和$\mathbf{u}_k^{(q)}$，所有这些量都可以通过其样本对应量进行一致估计。令

$$\widehat{\sigma}_0^2 = \frac{1}{\prod_{q=1}^{p}(d_q - r)} \sum_{1 \le i_q \le d_q, 1 \le q \le p} [\check{\Sigma}_{r+1}]_{i_1 \cdots i_p i_1 \cdots i_p}$$

and

$$\widehat{\sigma}_{\pi(k)}^2 = \langle \widehat{\Sigma}, \widehat{\mathscr{U}_k} \otimes \widehat{\mathscr{U}_k} \rangle - \widehat{\sigma}_0^2.$$

以下定理表明，如果我们用这些估计值替代线性形式的方差 $\langle \mathbf{u}_{\pi}^{(q)}{}_k()\mathbf{v}\rangle$，渐近正态性仍然成立：

定理 4.5。 设 $\mathscr{X}_1, \ldots, \mathscr{X}_n$ 是服从尖峰协方差模型 (8) 的独立观测，且 $p > 1$ 满足 $\mathscr{U}_k = \mathbf{u}_k^{(1)} \otimes \cdots \otimes \mathbf{u}_k^{(p)}$，并且 $\sigma_k > 0$。假设 r 和 $\sigma_1, \ldots, \sigma_r$均为固定值。存在一个置换 π： $[r] \to [r]$ 使得

(a) 若 $d = o(\sqrt{n})$，则

$$\frac{\sqrt{n}\left(\langle \tilde{\mathbf{u}}_k^{(q)}, \mathbf{v}\rangle - \langle \mathbf{u}_{\pi(k)}^{(q)}, \mathbf{v}\rangle\right)}{\sqrt{\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_{\pi(k)}^2} + \frac{\widehat{\sigma}_0^4}{\widehat{\sigma}_{\pi(k)}^4}}\left\|\mathcal{P}^{\perp}_{\tilde{\mathbf{u}}_{\pi(k)}^{(q)}}\mathbf{v}\right\|} \to_d N(0,1),$$

(b) 若 $d = o(n^{2/3})$，则

$$\frac{\sqrt{n}\left((1+b_k^{(q)})\langle \check{\mathbf{u}}_k^{(q)}, \mathbf{v}\rangle - \langle \mathbf{u}_{\pi(k)}^{(q)}, \mathbf{v}\rangle\right)}{\sqrt{\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_{\pi(k)}^2} + \frac{\widehat{\sigma}_0^4}{\widehat{\sigma}_{\pi(k)}^4}}\left\|\mathcal{P}^{\perp}_{\check{\mathbf{u}}_{\pi(k)}^{(q)}}\mathbf{v}\right\|} \to_d N(0,1),$$

其中 $b_k^{(q)}$ 由下式给出 (12)。

(c) 若 $d = o(n)$，则

$$\frac{\sqrt{n}\left((1+\widehat{b}_k^{(q)})\langle \check{\mathbf{u}}_k^{(q)}, \mathbf{v}\rangle - \langle \mathbf{u}_{\pi(k)}^{(q)}, \mathbf{v}\rangle\right)}{\sqrt{\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_{\pi(k)}^2} + \frac{\widehat{\sigma}_0^4}{\widehat{\sigma}_{\pi(k)}^4}}\left\|\mathcal{P}^{\perp}_{\check{\mathbf{u}}_{\pi(k)}^{(q)}}\mathbf{v}\right\|} \to_d N(0,1),$$

其中 $\widehat{b}_k^{(q)}$ 由下式给出 (13)。

定理 4.5 是斯卢茨基定理和定理 4.2-4.4的直接推论。它使我们能够对$\langle \mathbf{u}_{\pi(k)}^{(q)}, \mathbf{v}\rangle$进行推断或构造置信区间。例如，考虑检验以下假设：

$$H_0 : \langle \mathbf{u}_{\pi(k)}^{(q)}, \mathbf{v}\rangle = 0 \qquad vs \qquad H_a : \langle \mathbf{u}_{\pi(k)}^{(q)}, \mathbf{v}\rangle \neq 0,$$

当 $d = o(\sqrt{n})$ 时。我们可以当且仅当以下条件成立时拒绝 H_0

$$\left|\sqrt{n}\left(\langle \tilde{\mathbf{u}}_k^{(q)}, \mathbf{v}\rangle - \langle \mathbf{u}_{\pi(k)}^{(q)}, \mathbf{v}\rangle\right)\right| \geq z_{\alpha/2}\sqrt{\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_{\pi(k)}^2} + \frac{\widehat{\sigma}_0^4}{\widehat{\sigma}_{\pi(k)}^4}}\left\|\mathcal{P}^{\perp}_{\tilde{\mathbf{u}}_{\pi(k)}^{(q)}}\mathbf{v}\right\|,$$

其中，$z_{\alpha/2}$ 是标准正态分布的上侧 $\alpha/2$ 分位数。定理4.5保证这在渐近意义下是一个水平为α 的检验。同样，我们也可以为$\langle \mathbf{u}_{\pi(k)}^{(q)}, \mathbf{v}\rangle$构造（$1-\alpha$）置信区间：

$$\left(\langle \tilde{\mathbf{u}}_k^{(q)}, \mathbf{v}\rangle \pm \frac{z_{\alpha/2}}{\sqrt{n}}\sqrt{\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_{\pi(k)}^2} + \frac{\widehat{\sigma}_0^4}{\widehat{\sigma}_{\pi(k)}^4}}\left\|\mathcal{P}^{\perp}_{\tilde{\mathbf{u}}_{\pi(k)}^{(q)}}\mathbf{v}\right\|\right).$$

特别地，取 $\mathbf{v} \in \{\mathbf{e}_1, \ldots, \mathbf{e}^{d_q}\}$，即可利用上述公式推导 $\mathbf{u}_{\pi(k)}^{(q)}$ 坐标的置信区间。较大的 d 情形也可以相应地处理。

5 数值实验

为补充理论分析并进一步展示多路主成分分析的优点，我们开展了几组数值实验。

5.1 模拟研究

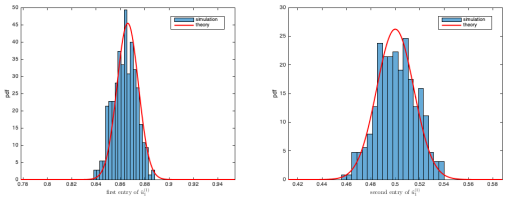
我们首先介绍一组模拟研究，以说明样本主成分在有限样本下的行为。这些实验专门用于评估偏差校正的作用，以及对偏离正态分布的稳健性。在本小节中，除非另有说明，样本均根据尖峰协方差模型（8）生成，其中$p = 2$，例如，每个 $\mathscr{X}_i$ 都是一个矩阵。由于两个模态可以互换，为简洁起见，我们只关注第一个模态 $q = 1$ 。我们还将尖峰数固定为 $r = 2$。在每种情形下，我们都设定奇异值 $\sigma_1 = \sigma_2$。换言之，在我们的每个例子中，由于重数的存在，传统主成分分析（采用串联）都无法识别主成分。如前所述，不失一般性且为简洁起见，我们对 $\check{\mathbf{u}}_1^{(1)}$ 和 $\check{\mathbf{u}}_2^{(1)}$ 重新排序，使得 $\sin\angle(\check{\mathbf{u}}_1^{(1)}, \mathbf{u}_1) \le \sin\angle(\check{\mathbf{u}}_2^{(1)}, \mathbf{u}_1)$。此外，每当 $\mathrm{h}\check{\mathbf{u}}_k^{(1)}, \mathbf{u}_k^{(1)}\,\mathrm{i} < 0$时，我们便将 $\check{\mathbf{u}}_k^{(1)}$ 替换为 $-\check{\mathbf{u}}_k^{(1)}$。对于低维设置，$\tilde{\mathbf{u}}_1^{(1)}$ 和 $\tilde{\mathbf{u}}_2^{(1)}$ 也采用类似处理。

在第一组实验中，我们考虑了一个包含 $d_1 = d_2 = 10$、$n = 200$、$\sigma_1 = \sigma_2 = 2$ 的低维设置，真实主成分为

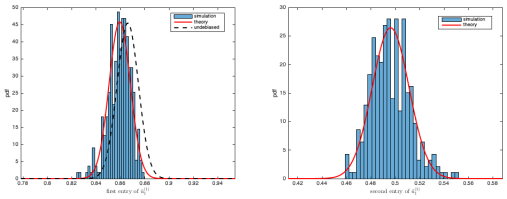
$$\begin{aligned}\mathbf{u}_1^{(1)} &= (\sqrt{3}/2, 1/2, 0, \ldots, 0)^\top, \quad \mathbf{u}_1^{(2)} = (1, 0, \ldots, 0)^\top,\\ \mathbf{u}_2^{(1)} &= (-1/2, \sqrt{3}/2, 0, \ldots, 0)^\top, \quad \mathbf{u}_2^{(2)} = (0, 1, 0, \ldots, 0)^\top.\end{aligned} \tag{14}$$

图 1a 报告了 $\tilde{\mathbf{u}}_1^{(1)}$ 的前两个（非零）元素基于 300 次模拟运行得到的直方图。直方图叠加了定理 4.1 中推导出的渐近分布。二者之间的一致性证实了当维数较低时渐近分布的准确性。

为展示偏差校正的必要性和有效性，我们将维度增加到 $d_1 = d_2 = 50$。相应地，我们设置 $n = 400$ 和 $\sigma_1 = \sigma_2 = 3$。我们又重复进行了 300 次实验，与之前一样，图1b报告了$\check{\mathbf{u}}_1^{(1)}$ 前两个分量的直方图，以及定理4.3导出的渐近分布，后者以红线绘出。叠加在第一个分量直方图上的黑色虚线，对应于定理4.1给出的未进行偏差校正时的渐近分布。显然



(a) $d_1 = d_2 = 10$



(b) $d_1 = d_2 = 50$

图 1：基于正态分布生成的数据进行多路主成分分析。

在这种情形下，去偏是必要的，而定理4.3中的偏差校正确实能更精确地近似有限样本分布。

接下来的一组模拟旨在探究我们的方法对偏离正态性的稳健性。为此，$\{\theta_k,\ k \in [r]\}$以及 $\mathscr{E}$的各个元素均独立地取自泊松分布(1) -1 (因此其均值仍为 0，方差仍为 1)。同样，我们设 $n = 400$且 $\sigma_1 = \sigma_2 = 3$。图 2a 和 2b 分别汇总了基于 300 次运行的结果，对应维度$d_1 = d_2 = 10$和 $d_1 = d_2 = 50$。我们将结果与定理 4.1 和 4.3给出的理论渐近分布叠加比较。结果在定性上与前一种设定下的结果相似。

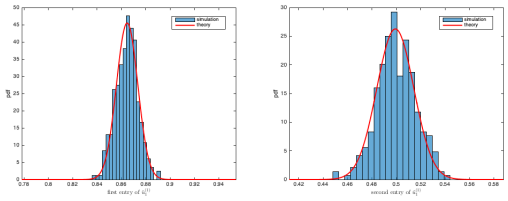
5.2 世界银行数据

现在我们来看一个真实数据示例——世界银行提供的开放源代码全球发展数据*。世界银行提供多种发展指标的年度国家层面数据。具体而言，我们将重点关注以下九项最常见且重要的经济和人口统计指标：

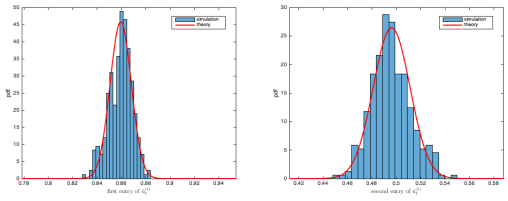
`GDP`：按购买力平价计算的国内生产总值（国内生产总值）；

`Import`：进口量指数（ 2000=100年）；

*https://data.worldbank.org/



(a) $d_1 = d_2 = 10$



(b) $d_1 = d_2 = 50$

图 2：由泊松分布生成的数据的多路主成分分析。

`Export`：出口量指数（ 2000=100年）；

`CO2`：二氧化碳排放量，单位为千吨；

`CPI`：消费者价格指数（ 2010 = 100 年）；

`Life Span`：出生时预期寿命；

`Urban Population`：城镇人口占总人口的百分比；

`Tourism`：国际入境游客人数；

`Birth Rate`：粗出生率（每 1,000 人）。

这些指标的年度数据均有记录，我们重点分析 2000 年至 2018 年的数据，因为在此范围之外缺失的数据相当多。我们还剔除了缺失数据超过 5% 的国家，最终纳入分析的国家总数为 160 个。

这些指标均为正值，但量级相差悬殊。因此，首先对其进行对数变换。随后，对每个经过对数变换的指标进行标准化，使其在所有国家中的均值为 0，平均绝对偏差为 1。与使用方差进行标准化相比，使用平均绝对偏差能够在存在离群观测时进行更稳健的分析。将国家 k 在时间 t 的结果指标记为 $\mathbf{X}_{k,t,i}$ i 。仍有少量缺失值，并且

为方便起见，在我们的分析中将这些值替换为 0。维度为$160 \times 19 \times 9$的数据张量 $\mathbf{X}$。每个正面切片

$$\mathscr{X}_k = \mathbf{X}_{k,\cdot,\cdot}$$

对应一个国家，是一个 19×9 矩阵。请注意，其环境维度为 $19 \times 9 =$171，大于国家数量，因此使用串联方式应用传统主成分分析会存在问题。考虑到多路结构，我们可以考察以下形式的多路主成分

$$\mathscr{U}_k = \mathbf{u}_k^{(1)} \otimes \mathbf{u}_k^{(2)} \in \mathbb{R}^{19\times 9}.$$

这些主成分具有明确的含义：每个 $\mathscr{U}_k$ 都代表一种共同的发展模式，其中$\mathbf{u}_k^{(1)}$ 是相应的共同时间趋势，而$\mathbf{u}_k^{(2)}$ 是相应的共同变动模式。

图3绘出了沿两个模态估计得到的首个主成分，即 $\check{}\mathbf{u}_1^{(1)}$ 和 $\check{}\mathbf{u}_k^{(2)}$，以及各自坐标的95%置信区间。它是迄今为止最显著的成分，解释了总变异的57.6%。从时间成分也可以明显看出，第一主成分描述了大致恒定的增长趋势。唯一出现下降的年份是2008年，当时全球金融危机爆发。相应地，除出生率对应的条目外，$\check{}\mathbf{u}_k^{(2)}$ 中的其他条目均为正值。这表明，在此期间整体经济有所发展，而出生率则有所下降。

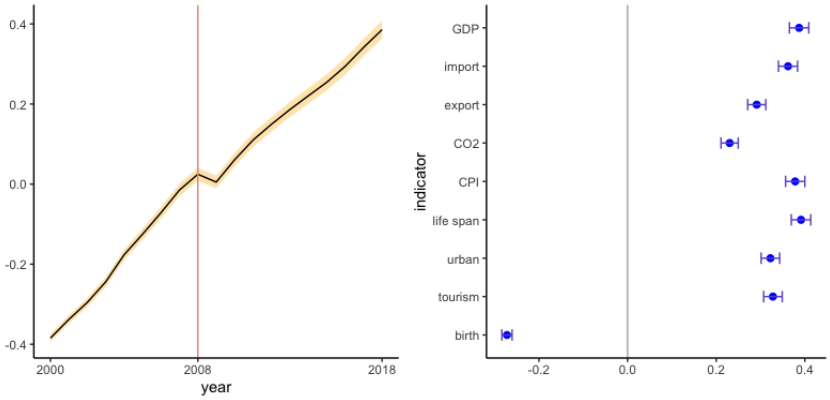


图3：第一个主成分，即总体经济发展：$\mathbf{u}_1^{(1)}$ 和$\mathbf{u}_1^{(2)}$，并绘制了95%的置信区间。

同样，图4展示了两个模态中的第二个多路主成分及其95%置信带。该主成分捕捉到了2008年发展方向的变化。

具体而言，消费者价格指数、寿命、城镇人口和旅游业在2008年之前稳步下降，但在金融危机后转而上升。相比之下，国内生产总值、进口、出口、二氧化碳排放和出生率则呈现相反的变化模式。对此有许多合理的解释。一种可能是，自2008年以来，大多数主要经济体实施的量化宽松政策推动了国内市场增长，进而改善了生活质量指标。另一种可能是，2008年后不平等加剧，这也是量化宽松等因素所致（例如，Montecino和Epstein，2015），反过来又导致上层和中上阶层的生活质量提高。旅游业指标指国际入境游客人数，这一群体很可能主要由中上阶层及以上人群推动。研究还指出，美国预期寿命持续增长，主要是由富裕人群推动的（例如，Chetty等人，2016）。

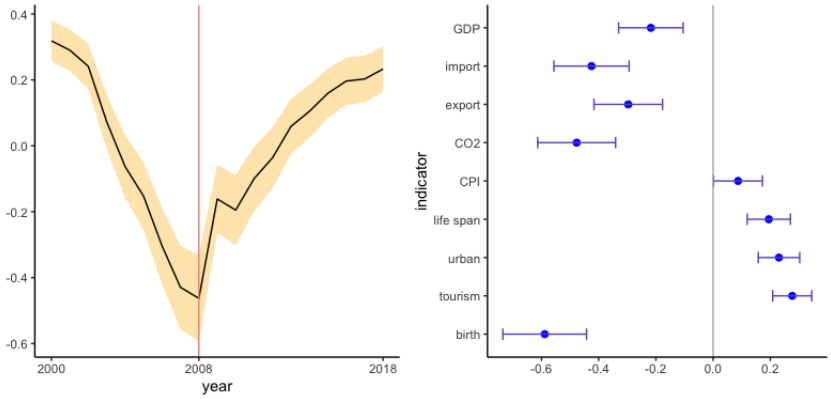


图4：第二主成分，生活质量：$\mathbf{u}_2^{(1)}$ 和$\mathbf{u}_2^{(2)}$，绘有95%置信区间。

最后，图5 展示了第三个多路主成分。随着信号变弱，我们开始看到置信区间明显变宽。事实上，只有2008年左右的时期与零显著不同；同样，与寿命、城市化和旅游业相对应的条目在统计上也不显著。这表明这些模式可能聚焦于2008年金融危机的影响：危机立即引发了经济衰退，但不久后便有所恢复。

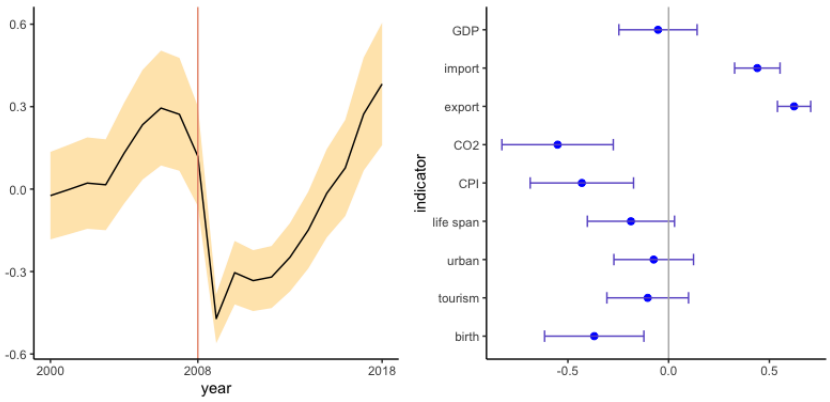


图5：第三个主成分，国际贸易：$\mathbf{u}_3^{(1)}$ 和 $\mathbf{u}_3^{(2)}$ ，附带95%置信区间。

5.3 纽约市自行车租赁数据

我们考虑的另一个数据示例是Citibike行程数据[†]。具体而言，我们在分析中使用了2018年1月1日至2019年12月31日期间所有工作日（共522天）从曼哈顿出发且持续至少60秒的Citibike行程。在此期间，曼哈顿有35个邮政编码区域至少设有一个Citibike站点。行程总数为29,515,527次，我们构建了一个维度为$\mathbf{Y}$的数据张量 $522 \times 24 \times 35$ 其中Y_{kij} 表示从 i日的 k时开始、起点位于 j邮政编码区域的行程数量。

不同邮政编码区域的计数数量差异悬殊，而且两年期间的总计数也呈现出明显的季节性趋势。为便于分析，我们对每个邮政编码区域 j 在每天 k 的计数进行标准化，使其均值为0，平均绝对偏差为1。与前面的示例一样，直接应用传统主成分分析可能会产生误导，因为每日观测的环境维度为$24 \times 35 = 840$，大于天数。尽管如此，考虑以下形式的多路主成分仍有帮助：

$$\mathscr{U}_k = \mathbf{u}_k^{(1)} \otimes \mathbf{u}_k^{(2)} \in \mathbb{R}^{24 \times 35},$$

其中，$\mathbf{u}_k^{(1)}$ 捕捉自行车租赁的日内时间效应，而$\mathbf{u}_k^{(2)}$ 则捕捉位置模式。

图 6 展示了第一个多路主成分。空间模式清楚地表明，这代表了覆盖整个曼哈顿的总体模式，其中$\mathbf{u}_k^{(2)}$ 的全部 35 个条目均被估计为正值

[†]https://ride.citibikenyc.com/system-data

。时间模式表明，自行车租赁与高峰时段密切吻合，早间和下午高峰时段各有一个峰值。市中心的空白区域是邮政编码 10006，那块较大的矩形空白区域是中央公园，其下方较小的空白区域是邮政编码 10020，而中央公园以北的空白区域涵盖邮政编码 10030 和 10031。在记录期间，这些区域均未设立 Citibike站点。

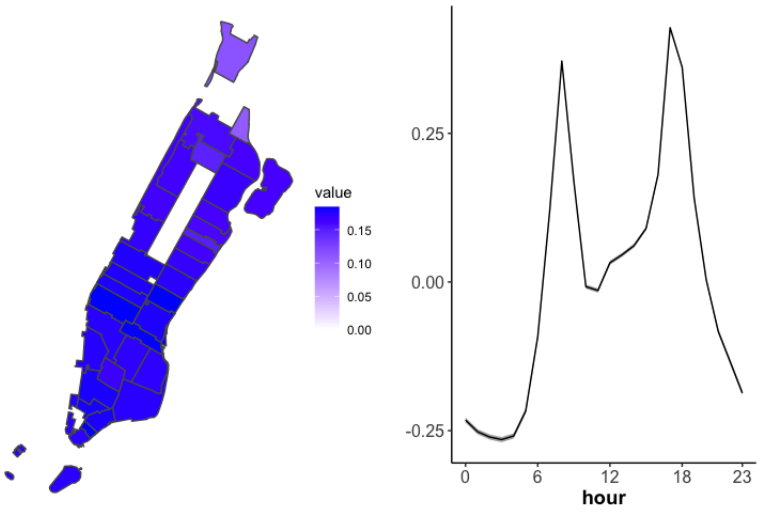


图 6：第一个主成分：整体模式。

如图 7 所示，第二个主成分揭示了不同街区之间租赁模式的差异。第一个主成分显示，早晨和下午高峰时段的租赁活动都会增加；而第二个主成分则反映了早晚租赁模式的差异：晚间高峰时段出现正峰值，早晨高峰时段则出现负峰值。因此，载荷为正的街区晚间租赁量可能多于早晨。这些街区包括市中心的金融区、下曼哈顿和中城，大致对应曼哈顿的商业区。另一方面，载荷为负的邮政编码对应的区域大多是曼哈顿的住宅区，包括东村、上西区和上东区。

图 8 展示了第三主成分。其时间模式在下午高峰时段呈现一个狭窄而高耸的峰值，这表明该主成分捕捉到了这段时间细微的空间差异。具体来说，取值较大的邮政编码区域（紫色）包括华尔街周边区域（下曼哈顿的一小块紫色区域）、

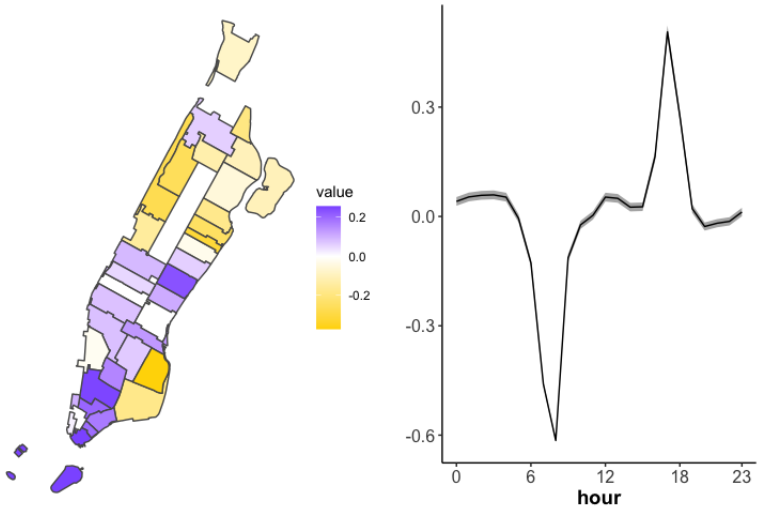


图7：第二主成分：高峰时段的差异。

中央车站周边区域，以及上东区的一处区域。该模式中取值为负的邮政编码区域包括苏荷区、格林威治村和哈莱姆区周边区域。

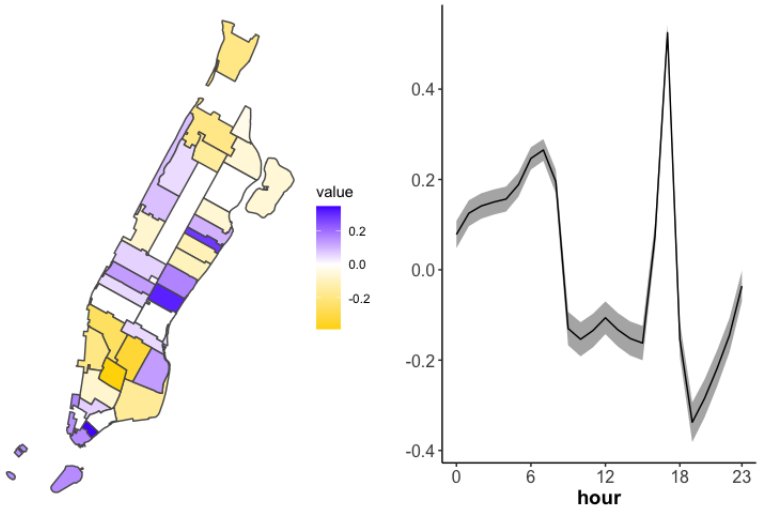


图8：第三主成分：下午高峰时段的细节。

6 总结

本文研究每个观测均为矩阵或更一般的多维数组时的主成分分析。我们探讨如何提取多路主成分，并研究其

统计性质。除了显而易见的效率提升和可解释性增强之外，我们的分析还为多路主成分分析的运行特性及其方法论意义提供了若干新见解。

首先，我们证明无需满足特征值间隙要求即可估计多路主成分。具体而言，在尖峰协方差模型下，我们建立了样本多路主成分的收敛速率。特别是，只要信噪比 $\sigma_k/\sigma_0 \gg \max\{d/n,(d/n)^{1/4}\}$ ，它们就是一致的，其中 d 是一个模态的维数。或许更有意思的是，我们证明样本多路主成分彼此渐近独立，至少当维数 $d = o(\sqrt{n})$时如此。在更高维度下，样本主成分可能存在偏差；通过样本拆分可以校正该偏差，从而得到多路主成分的渐近正态估计，进而构造主成分线性形式的置信区间或进行假设检验。

我们通过数值实验（包括模拟实验和真实数据实验）补充了理论研究。特别是，将我们的方法应用于两个真实世界的数据集时，可以得出有意义的发现，进一步展示了我们方法的优点。

参考文献

Animashree Anandkumar、Rong Ge、Daniel Hsu、Sham M Kakade 和 Matus Telgarsky。用于学习潜变量模型的张量分解。机器学习研究期刊，15：2773–2832，2014。

T. W. Anderson。《多元统计分析导论》。Wiley，纽约州纽约市，第二版，1984。

Arnab Auddyand Ming Yuan。近似正交可分解张量的扰动界。*arXiv* 预印本 *arXiv:2007.09024*，2020。

ZhidongBai 和 Jack W Silverstein。大维随机矩阵的谱分析，第20卷。施普林格，2010。

Zhidong Bai 和 Jianfeng Yao。《广义尖峰总体模型中的样本特征值》。《多元分析杂志》，106：167–177，2012 年。

JinhoBaik 和 Jack WSilverstein。《尖峰总体模型的大型样本协方差矩阵的特征值》。《多元分析杂志》，97(6)：1382–1408，2006 年。

Florent Benaych-Georges 和 Raj Rao Nadakuditi。大型随机矩阵有限秩扰动的特征值和特征向量。《数学进展》，227 (1):494–521，2011。

Xuan Bi、Xiwei Tang、Yubai Yuan、Yanqing Zhang 和 Annie Qu。统计学中的张量。《统计学及其应用年度评论》，8:345–368，2021。

Aharon Birnbaum、Iain M Johnstone、Boaz Nadler 和 Debashis Paul。含噪高维数据下稀疏主成分分析的极小极大界。《统计年刊》，41(3):1055，2013。

Elynn Y Chen、Jianqing Fan 和 EllenLi。高维矩阵变元因子模型的统计推断。*arXiv* 预印本 *arXiv:2001.01890*，2020a。

Elynn Y Chen、Dong Xia、Chencheng Cai 和 Jianqing Fan。通过迭代投影奇异值分解进行半参数张量因子分析。*arXiv* 预印本 *arXiv:2007.02404*，2020b。

Rong Chen、Dan Yang 和 Cun-Hui Zhang。《高维张量时间序列的因子模型》。《美国统计学会杂志》，第 1–23 页，2021 年。

Raj Chetty、Michael Stepner、Sarah Abraham、Shelby Lin、Benjamin Scuderi、Nicholas Turner、Augustin Bergeron 和 David Cutler。《2001—2014年美国收入与预期寿命之间的关联》。《美国医学会杂志》，315(16):1750–1766，2016 年。

Andrzej Cichocki、Danilo Mandic、Lieven De Lathauwer、Guoxu Zhou、Qibin Zhao、Cesar Caiafa 和 Huy Anh Phan。《面向信号处理应用的张量分解：从双向到多向成分分析》。《*IEEE*信号处理杂志》，32(2): 145–163，2015 年。

Lieven De Lathauwer、Bart De Moor 和 Joos Vandewalle。《多线性奇异值分解》。*SIAM* 矩阵分析与应用杂志，21(4):1253–1278，2000 年。

Shmuel Friedland. 实对称张量的最佳秩一近似可取为对称形式。《中国数学前沿》，8(1):19–40，2013。

Wolfgang Hackbusch. 张量空间与数值张量微积分，第 42 卷。施普林格，2012。Rungang Han、Rebecca Willett 和 Anru R Zhang. 广义张量估计的最优统计与计算框架。《统计年鉴》，50(1):1–29，2022。Yuefeng Han、Rong Chen、Dan Yang 和 Cun-Hui Zhang. 通过迭代投影进行张量因子模型估计。*arXiv* 预印本 *arXiv:2006.02611*，2020。

Richard AHarshman 和 MargaretE Lundy。三路因子分析与多维尺度分析的PARAFAC模型。《多模态数据分析研究方法》，46:122–215，1984。

Samuel B Hopkins、Jonathan Shi 和 David Steurer。通过平方和证明进行张量主成分分析。载于《学习理论会议》，第 956–1006 页，2015。

Majid Janzamin、Rong Ge、Jean Kossaifi、Anima Anandkumar 等。矩阵与张量上的谱学习。《机器学习基础与趋势》，12(5-6):393–536，2019。

Iain M Johnstone。主成分分析中最大特征值的分布。《统计年刊》，第 295–327 页，2001。

Iain M Johnstone 和 Arthur Yu Lu. 高维下主成分分析的一致性与稀疏性。《美国统计学会杂志》，104(486): 682–693，2009。I. Jolliffe. 主成分分析。施普林格，2002。Sungkyu Jung 和 J Stephen Marron。高维小样本情境下主成分分析的一致性。《统计年鉴》，37(6B):4104–4130，2009。Tamara G Kolda。正交张量分解。*SIAM* 矩阵分析与应用杂志，23(1):243–255，2001。Vladimir Koltchinskii 和 Karim Lounici。样本协方差谱投影算子的渐近性质与集中界。*arXiv* 预印本 *arXiv:1408.4643*，2014。

Vladimir Koltchinskii、Karim Lounici 等。样本协方差算子的集中不等式与矩界。*Bernoulli*，23(1):110–133，2017。

Vladimir Koltchinskii、Matthias L¨offler 和 Richard Nickl。主成分线性泛函的有效估计。《统计年鉴》，48(1):464–490，2020。

Hui Kong、Lei Wang、Eam Khwang Teoh、Xuchun Li、Jian-Gang Wang 和 Ronda Venkateswarlu。用于人脸图像表示与识别的广义二维主成分分析。《神经网络》，18(5-6):585–594，2005。

Pieter M Kroonenberg。应用多向数据分析。约翰·威立父子出版公司，2008年。

Pieter M Kroonenberg 和 Jan De Leeuw。通过交替最小二乘算法对三模态数据进行主成分分析。《心理测量学》，45(1):69–97，1980。

Seunggeun Lee、Fei Zou 和 Fred A Wright。高维情形下主成分得分的收敛与预测。《统计年刊》，38(6):3605，2010。

Xuelong Li、Yanwei Pang 和 Yuan Yuan。基于L1范数的二维主成分分析。《*IEEE*系统、人类与控制论汇刊*B*辑（控制论）》，40(4):1170–1175，2010。

Tianqi Liu、Ming Yuan 和 Hongyu Zhao。通过低秩张量分解表征人脑时空转录组。*arXiv* 预印本 *arXiv:1702.07449*，2017年。

Haiping Lu、Konstantinos N Plataniotis 和 Anastasios N Venetsanopoulos。用于识别的张量对象多线性主成分分析。载于第*18*届国际模式识别会议（*ICPR'06*），第2卷，第776–779页。IEEE，2006年。

Haiping Lu、Konstantinos N Plataniotis 和 Anastasios N Venetsanopoulos。多线性主成分分析：张量对象的多线性主成分分析。《*IEEE*神经网络汇刊》，19(1):18–39，2008年。

Haiping Lu、Konstantinos N Plataniotis 和 Anastasios N Venetsanopoulos。张量数据多线性子空间学习综述。《模式识别》，44(7):1540–1551，2011年。Juan Montecino 和 Gerald Epstein。量化宽松是否加剧了收入不平等？新经济思想研究所工作论文系列，(28)，2015。

Boaz Nadler。主成分分析的有限样本近似结果：一种矩阵扰动方法。《统计年鉴》，36(6):2791–2817，2008年。Debashis Paul。大型维数尖峰协方差模型中样本特征结构的渐近性质。《中华统计学志》，第1617–1642页，2007年。Emile Richard 和 Andrea Montanari。张量主成分分析的统计模型。载于神经信息处理系统进展，第2897–2905页，2014。

Dan Shen、Haipeng Shen、Hongtu Zhu 和 J S Marron。临界样本特征方向中令人惊讶的渐近锥形结构。*arXiv* 预印本 *arXiv:1303.6171*，2013。

Y-H Taguchi。基于张量分解和主成分分析的无监督特征提取，应用于具有多个阶级的社会性昆虫大脑中的基因表达和甲基化谱。《*BMC*生物信息学》，19(4):99，2018。

M Alex O Vasilescu 和 Demetri Terzopoulos。图像集的多线性分析：张量脸。载于欧洲计算机视觉会议，第 447–460 页。施普林格，2002。

Roman Vershynin。随机矩阵非渐近分析导论。*arXiv* 预印本 *arXiv:1011.3027*，2010。

Weichen Wang 和 Jianqing Fan。高维尖峰协方差的经验特征结构的渐近性质。《统计年刊》，45(3):1342，2017。

Dong Xia、Anru R Zhang 和 Yuchen Zhou。低秩张量的推断——无需去偏差。*arXiv* 预印本 *arXiv:2012.14844*，2020 年。

Jian Yang、David Zhang、Alejandro F Frangi 和 Jing-yu Yang。二维主成分分析：一种用于基于外观的人脸表示与识别的新方法。《*IEEE*模式分析与机器智能汇刊》，26(1)：131–137，2004 年。

Anru Zhang 和 Dong Xia。张量奇异值分解：统计与计算极限。*IEEE* 信息论汇刊，64(11)：7311–7338，2018 年。

Daoqiang Zhang 和 Zhi-Hua Zhou。（2d）2pca：用于高效人脸表示与识别的双向二维主成分分析。《神经计算》，69(1-3):224–231，2005。

Tong Zhang 和 Gene H Golub。高阶张量的秩一近似。*SIAM* 矩阵分析与应用杂志，23(2):534–550，2001。

A 记号与初步界

记 $a \vee b := \max\{a, b\}$ 和 $a \wedge b := \min\{a, b\}$。对于正整数 n，令 $[n] := \{1, 2, \ldots, n\}$。对于向量 $\mathbf{x} \in \mathbb{R}^d$，记 $\|\mathbf{x}\|$ 为其 ℓ_2-范数，$\|\mathbf{x}\|_1$ 为其 ℓ_1-范数，$\|\mathbf{x}\|_\infty = \max_i |x_i|$ 为其 ℓ_∞-范数。对于两个实数序列 $\{a_n\}$ 和 $\{b_n\}$，若 $\exists C, \exists M$，使得 $\forall n > M$，$|a_n| \leq C|b_n|$，则记为 $a_n = O(b_n)$。若 $\lim_{n\to\infty} a_n/b_n = 0$，则记为 $a_n = o(b_n)$。对于两个实值随机变量序列 X_n 和 Y_n，若 $X_n = R_n Y_n$ 且 R_n 一致紧，则记为$X_n = O_p(Y_n)$。若 $X_n = R_n Y_n$ 且 $R_n \xrightarrow{p} 0$，则记为 $X_n = o_p(Y_n)$。对于 $\mathbb{R}^d$ 的线性子空间 U，记 P_U 和 $P_U^\perp$分别为到U 及其正交补 $U^\perp$的正交投影。对于非零向量 $u \in \mathbb{R}^d$，记$P_u := P_{\mathrm{span}\{u\}}$ 和 $P_u^\perp := P_{\mathrm{span}\{u\}}^\perp$。

对于一个阶数为 k、尺寸为 $\mathscr{T} \in \mathbb{R}^{d_1 \times d_2 \times \ldots \times d_k}$ 的张量，定义其张量算子范数 为:

$$\|\mathscr{T}\| := \sup_{\mathbf{u}_j \in \mathbb{R}^{d_j}, \|\mathbf{u}_j\|=1} \mathscr{T}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k). \tag{15}$$

具体而言，当 $k = 2$ 使得 $\mathscr{T} \in \mathbb{R}^{d_1 \times d_2}$ 是一个矩阵时，$\|\mathscr{T}\|$ 是 $\mathscr{T}$ 的矩阵谱范数。对于张量 $\mathscr{T} \in \mathbb{R}^{d_1 \times d_2 \times \ldots \times d_k}$，记

$$\|\mathscr{T}\|_{\max} = \max_{i_1, \ldots, i_k} |\mathscr{T}_{i_1, \ldots, i_k}|$$

为其 ℓ_∞-范数。

稍微滥用一下记号，$\mathscr{T} \in \mathbb{R}^{d_1 \times d_2 \times \ldots \times d_k}$ 与向量$\mathbf{a} \in \mathbb{R}^{d_q}$ 的模态 q 乘积记为 $\mathscr{T} \times_q \mathbf{a}$，定义为一个阶数为 $(k-1)$、尺寸为 $d_1 \times \ldots d_{q-1} \times d_{q+1} \cdots \times d_k$ 的张量，其元素为

$$[\mathscr{T} \times_q \mathbf{a}]_{i_1 \ldots i_{q-1} i_{q+1} \ldots i_k} = \sum_{i_q=1}^{d_q} \mathscr{T}_{i_1 \ldots i_q \ldots i_k} \mathbf{a}_{i_q}.$$

记

$$\widehat{\Sigma}_\theta = \frac{1}{n}\sum_{i=1}^n \theta_i \otimes \theta_i, \qquad \widehat{\Sigma}_{\mathscr{E}} = \frac{1}{n}\sum_{i=1}^n \mathscr{E}_i \otimes \mathscr{E}_i,$$

and

$$\widehat{\Sigma}_{\theta,\mathscr{E}} = \frac{1}{n}\sum_{i=1}^n \theta_i \otimes \mathscr{E}_i,$$

θ、 $\mathscr{E}$ 及它们之间的样本协方差矩阵。相应地，用 Σ_θ、 $\Sigma_{\mathscr{E}}$ 和$\Sigma_{\theta,\,\mathscr{E}}$ 表示它们总体中的对应矩阵。显然$\Sigma_{\theta,\,\mathscr{E}} = 0$。还需回顾

that

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \mathscr{X}_i \otimes \mathscr{X}_i.$$

and

$$\Sigma = \sum_{l=1}^{r} \sigma_l \mathbf{u}_l^{(1)} \otimes \cdots \otimes \mathbf{u}_l^{(p)} \otimes \mathbf{u}_l^{(1)} \otimes \cdots \otimes \mathbf{u}_l^{(p)} + \sigma_0^2 \mathscr{I}$$

分别是 $\mathscr{X}$的样本协方差矩阵和总体协方差矩阵。

证明依赖于以下技术性引理。

引理 1. 存在一个数值常数 $C > 0$，使得对于任意 $t \geq 1$,

$$\|\widehat{\Sigma} - \Sigma\| \leq C(\sigma_1^2 + \sigma_0^2) \max\left\{\sqrt{\frac{d}{n}}, \frac{d}{n}, \sqrt{\frac{t}{n}}, \frac{t}{n}\right\},$$

$$\|\widehat{\Sigma}_{\theta,\mathscr{E}}\| \leq C\sigma_0 \max\left\{\sqrt{\frac{d}{n}}, \frac{d}{n}, \sqrt{\frac{t}{n}}, \frac{t}{n}\right\},$$

$$\|\widehat{\Sigma}_{\mathscr{E}} - \Sigma_{\mathscr{E}}\| \leq C\sigma_0^2 \max\left\{\sqrt{\frac{d}{n}}, \frac{d}{n}, \sqrt{\frac{t}{n}}, \frac{t}{n}\right\},$$

and

$$\|\widehat{\Sigma}_{\theta} - \Sigma_{\theta}\| \leq C \max\left\{\sqrt{\frac{r}{n}}, \frac{r}{n}, \sqrt{\frac{t}{n}}, \frac{t}{n}\right\},$$

其概率至少为 $1 - e^{-t}$。

注意，我们将用 C 表示一个在不同位置可能取不同值的常数。我们还将使用以下界：

Lemma 2. 存在一个数值常数 $C > 0$，使得对于任意 $t >$ 0,

$$\|\widehat{\Sigma}_{\theta} - \Sigma_{\theta}\|_{\max} \leq C \max\left\{\sqrt{\frac{\log r}{n}}, \frac{\log r}{n}, \sqrt{\frac{t}{n}}, \frac{t}{n}\right\},$$

$$\max_{1\leq l_1,l_2\leq r} \left|\widehat{\Sigma}_{\mathscr{E},\theta}(\mathbf{u}_{l_1}^{(1)}, \mathbf{u}_{l_2}^{(2)}, \ldots, \mathbf{u}_{l_2}^{(p)}, \mathbf{e}_{l_2})\right| \leq C\sigma_0 \max\left\{\sqrt{\frac{\log r}{n}}, \frac{\log r}{n}, \sqrt{\frac{t}{n}}, \frac{t}{n}\right\}$$

其中$\mathbf{e}_{l_2}$ 是 $\mathbb{R}^r$的第 l_2 个标准基，且

$$\max_{1\leq l\leq r} \left|(\widehat{\Sigma}_{\mathscr{E}} - \Sigma_{\mathscr{E}})(\mathbf{u}_l^{(1)}, \ldots, \mathbf{u}_l^{(p)}, \mathbf{u}_l^{(1)}, \ldots, \mathbf{u}_l^{(p)})\right| \leq C\sigma_0^2 \max\left\{\sqrt{\frac{\log r}{n}}, \frac{\log r}{n}, \sqrt{\frac{t}{n}}, \frac{t}{n}\right\},$$

其概率至少为 $1 - e^{-t}$。

这两个引理广为人知，可直接由并集界和 χ^2 尾部界得到。例如，参见 Vershynin (2010)。

B 定理3.1和 3.2的证明

定理 3.1 可由定理 3.2直接推出，因此只需证明后者。为简洁起见，我们重点讨论 $d \le n$ 且 r 以 n发散的情形。记 $\mathcal{E}$ 为以下事件:

$$\|\widehat{\Sigma}_\theta - \Sigma_\theta\| \le C\sqrt{\frac{r}{n}}, \quad \text{and} \quad \sigma_0^{-2}\|\widehat{\Sigma}_{\mathscr{E}} - \Sigma_{\mathscr{E}}\|, \sigma_0^{-1}\|\widehat{\Sigma}_{\theta,\mathscr{E}}\| \le C\sqrt{\frac{d}{n}}$$

and

$$\|\widehat{\Sigma}_\theta - \Sigma_\theta\|_{\max}, \ \sigma_0^{-1} \max_{1\le l_1,l_2\le r} \left|\widehat{\Sigma}_{\mathscr{E},\theta}(\mathbf{u}_{l_1}^{(1)}, \mathbf{u}_{l_2}^{(2)}, \ldots, \mathbf{u}_{l_2}^{(p)}, \mathbf{e}_{l_2})\right| \le C\sqrt{\frac{\log r}{n}}$$

根据引理1 和 2，$\mathcal{E}$ 以趋于 1 的概率成立。只需在条件于事件 $\mathcal{E}$的情形下进行。

如前所述，由于我们不假设存在特征值间隙，且 σ_k 可能彼此不相同，第 k 个样本主成分可能并不对应于第 k 个总体主成分。尽管如此，我们可以按如下方式将样本主成分与总体主成分匹配。定义

$$\pi(1) = \text{argmax}_{1\le l\le r}\left\{\sigma_l^2 \left|\prod_{q=1}^{p}\langle \mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_1^{(q)}\rangle\right|\right\},$$

对于 $k > 1$,

$$\pi(k) := \text{argmax}_{l\notin\pi([k-1])}\left\{\sigma_l^2 \left|\prod_{q=1}^{p}\langle \mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_k^{(q)}\rangle\right|\right\}.$$

目标是证明，以高概率,

$$\eta_k := \max_{1\le q\le p} \sin\angle(\mathbf{u}_{\pi(k)}^{(q)}, \widehat{\mathbf{u}}_k^{(q)}) \le C\left(\frac{\sigma_0}{\sigma_{\pi(k)}} + \frac{\sigma_0^2}{\sigma_{\pi(k)}^2}\right)\max\left\{\sqrt{\frac{d}{n}}, \frac{d}{n}\right\} =: \delta_k, \tag{16}$$

对 $k = 1$、.....、 r 均成立。我们的证明通过对 k 作归纳法进行。为便于归纳，我们还将证明

$$\begin{aligned}\tilde{\eta}_k &:= \max_{1\le q\le p}\max_{l\notin\pi([k])}\langle \mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_k^{(q)}\rangle \\ &\le C\left(\frac{\sigma_0}{\sigma_{\pi(k)}} + \frac{\sigma_0^2}{\sigma_{\pi(k)}^2}\right)^2 \max\left\{\frac{d}{n}, \frac{d^2}{n^2}\right\} + C\left(\frac{\sigma_0}{\sigma_{\pi(k)}} + \frac{\sigma_0^2}{\sigma_{\pi(k)}^2}\right)\sqrt{\frac{\log r}{n}} \\ &=: \tilde{\delta}_k. \end{aligned} \tag{17}$$

除了（16）和（17）之外，我们还将证明

$$\sigma_{\pi(k)}^2 \ge \max_{l\notin\pi([k])} \sigma_l^2(1 - C\delta_k^2). \tag{18}$$

这立即意味着

$$\sum_{l=1}^{k} \tilde{\delta}_k^2 \le C\delta_k^2 \quad \text{and} \quad \max_{1\le l\le k}\{\sigma_{\pi(l)}\delta_l\} \le C\sigma_{\pi(k)}\delta_k,$$

通过将常数 c_0 在（10）中取足够小。我们将反复使用这些界。

如前所述，我们将对 k进行归纳。具体而言，我们记 $\delta_0 = \tilde{\delta}_0 = 0$ 因此，当 $k = 0$时，基础情形显然成立。现在假设归纳假设（16）和（17）对 $\mathbf{1}$，...，$k-1$成立。我们要证明它们对 k仍然成立。论证的整体思路与基础情形类似，但还需应对额外的挑战：控制 $\hat{\mathbf{u}}_l^{(q)}$ 在 $1 \le l < k$处的估计误差所产生的影响。

记 $\hat{\mathcal{P}}^{(q)}$ 为投影到由$\{\hat{\mathbf{u}}_1^{(q)}$，...，$\hat{\mathbf{u}}_{k-1}^{(q)}\}$张成的线性空间上的投影矩阵，其中 $q \in [p]$。注意，当 $k = 1$, $\hat{\mathcal{P}}^{(q)} = \mathbf{0}_{d\times d}$ 时。于是

$$\begin{aligned}(\hat{\mathbf{u}}_k^{(1)}, \dots, \hat{\mathbf{u}}_k^{(p)}) &= \operatorname{argmax}_{\substack{\|\mathbf{w}^{(q)}\|\le 1, \langle \mathbf{w}^{(q)}, \hat{\mathbf{u}}_l^{(q)}\rangle = 0, \\ \forall l\le k-1, q\in[p]}} \widehat{\Sigma}(\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(p)}, \mathbf{w}^{(1)}, \dots, \mathbf{w}^{(p)}) \\ &= \operatorname{argmax}_{\|\mathbf{w}^{(q)}\|=1, \forall q\in[p]} \widehat{\Sigma}(\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w}^{(1)}, \dots, \widehat{\mathcal{P}}_\perp^{(p)}\mathbf{w}^{(p)}, \widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w}^{(1)}, \dots, \widehat{\mathcal{P}}_\perp^{(p)}\mathbf{w}^{(p)})\end{aligned}$$

其中 $\hat{\mathcal{P}}_\perp^{(q)} = I - \hat{\mathcal{P}}^{(q)}$。注意到

$$\widehat{\mathcal{P}}_\perp^{(q)}\tilde{\Sigma}(\hat{\mathbf{u}}_k^{(1)}, \dots, \hat{\mathbf{u}}_k^{(q-1)}, \cdot, \hat{\mathbf{u}}_k^{(q+1)}, \dots, \hat{\mathbf{u}}_k^{(p)}, \hat{\mathbf{u}}_k^{(1)}, \dots, \hat{\mathbf{u}}_k^{(p)}) \propto \hat{\mathbf{u}}_k^{(q)},$$

其中 $\tilde{\Sigma} = \widehat{\Sigma} - \sigma_0^2\mathscr{I}$。这意味着

$$\langle \hat{\mathbf{u}}_k^{(q)}, \mathbf{w}\rangle = \frac{\tilde{\Sigma}(\hat{\mathbf{u}}_k^{(1)}, \dots, \hat{\mathbf{u}}_k^{(q-1)}, \widehat{\mathcal{P}}_\perp^{(q)}\mathbf{w}, \hat{\mathbf{u}}_k^{(q+1)}, \dots, \hat{\mathbf{u}}_k^{(p)}, \hat{\mathbf{u}}_k^{(1)}, \dots, \hat{\mathbf{u}}_k^{(p)})}{\tilde{\Sigma}(\hat{\mathbf{u}}_k^{(1)}, \hat{\mathbf{u}}_k^{(2)}, \dots, \hat{\mathbf{u}}_k^{(p)}, \hat{\mathbf{u}}_k^{(1)}, \dots, \hat{\mathbf{u}}_k^{(p)})}.$$

特别地，对于 $l \ne k$,

$$\sin\angle(\hat{\mathbf{u}}_k^{(q)}, \mathbf{u}_l^{(q)}) = \frac{\tilde{\Sigma}(\hat{\mathbf{u}}_k^{(1)}, \dots, \hat{\mathbf{u}}_k^{(q-1)}, \mathbf{u}_l^{(q)}, \hat{\mathbf{u}}_k^{(q+1)}, \dots, \hat{\mathbf{u}}_k^{(p)}, \hat{\mathbf{u}}_k^{(1)}, \dots, \hat{\mathbf{u}}_k^{(p)})}{\tilde{\Sigma}(\hat{\mathbf{u}}_k^{(1)}, \hat{\mathbf{u}}_k^{(2)}, \dots, \hat{\mathbf{u}}_k^{(p)}, \hat{\mathbf{u}}_k^{(1)}, \dots, \hat{\mathbf{u}}_k^{(p)})},$$

and

$$\sin\angle(\hat{\mathbf{u}}_k^{(q)}, \mathbf{u}_k^{(q)}) = \frac{\tilde{\Sigma}(\hat{\mathbf{u}}_k^{(1)}, \dots, \hat{\mathbf{u}}_k^{(q-1)}, \widehat{\mathcal{P}}_\perp^{(q)}(\hat{\mathbf{u}}_k^{(q)} - \langle \hat{\mathbf{u}}_k^{(q)}, \mathbf{u}_{\pi(k)}^{(q)}\rangle \mathbf{u}_{\pi(k)}^{(q)}), \hat{\mathbf{u}}_k^{(q+1)}, \dots, \hat{\mathbf{u}}_k^{(p)}, \hat{\mathbf{u}}_k^{(1)}, \dots, \hat{\mathbf{u}}_k^{(p)})}{\tilde{\Sigma}(\hat{\mathbf{u}}_k^{(1)}, \hat{\mathbf{u}}_k^{(2)}, \dots, \hat{\mathbf{u}}_k^{(p)}, \hat{\mathbf{u}}_k^{(1)}, \dots, \hat{\mathbf{u}}_k^{(p)})},$$

我们将推导分子的下界以及分母的上界。只需考虑 $q = 1$ 的情形。其他指标也可用完全相同的方法处理。

B.1 $\tilde{\Sigma}$（$\hat{\mathbf{u}}_k^{(1)}$，$\hat{\mathbf{u}}_k^{(2)}$，...，$\hat{\mathbf{u}}_k^{(p)}$，$\hat{\mathbf{u}}_k^{(1)}$，...，$\hat{\mathbf{u}}_k^{(p)}$）的下界

记为 $\tilde{\Sigma} = \Sigma - \sigma_0^2 \mathscr{I}$。注意到

$$\begin{aligned}
& \tilde{\Sigma}(\widehat{\mathbf{u}}_k^{(1)}, \widehat{\mathbf{u}}_k^{(2)}, \ldots, \widehat{\mathbf{u}}_k^{(p)}, \widehat{\mathbf{u}}_k^{(1)}, \ldots, \widehat{\mathbf{u}}_k^{(p)}) \\
\geq & \max_{l \notin \pi([k-1])} \tilde{\Sigma}(\widehat{\mathcal{P}}_\perp^{(1)} \mathbf{u}_l^{(1)}, \ldots, \widehat{\mathcal{P}}_\perp^{(p)} \mathbf{u}_l^{(p)}, \widehat{\mathcal{P}}_\perp^{(1)} \mathbf{u}_l^{(1)}, \ldots, \widehat{\mathcal{P}}_\perp^{(p)} \mathbf{u}_l^{(p)}) \\
\geq & \max_{l \notin \pi([k-1])} (\Sigma - \sigma_0^2 \mathscr{I}) \widehat{\mathcal{P}}_\perp^{(1)} \mathbf{u}_l^{(1)}, \ldots, \widehat{\mathcal{P}}_\perp^{(p)} \mathbf{u}_l^{(p)}, \widehat{\mathcal{P}}_\perp^{(1)} \mathbf{u}_l^{(1)}, \ldots, \widehat{\mathcal{P}}_\perp^{(p)} \mathbf{u}_l^{(p)}) \\
& - \sup_{\|\mathbf{w}^{(q)}\| \leq 1, 1 \leq q \leq p} \left| (\widehat{\Sigma} - \Sigma)(\widehat{\mathcal{P}}_\perp^{(1)} \mathbf{w}^{(1)}, \ldots, \widehat{\mathcal{P}}_\perp^{(p)} \mathbf{w}^{(p)}, \widehat{\mathcal{P}}_\perp^{(1)} \mathbf{w}^{(1)}, \ldots, \widehat{\mathcal{P}}_\perp^{(p)} \mathbf{w}^{(p)}) \right|. \qquad (19)
\end{aligned}$$

接下来，我们对最右端的两个项进行界定。

先看第一项。注意，对于任意 $l \notin \pi([k-1])$，

$$\begin{aligned}
& (\Sigma - \sigma_0^2 \mathscr{I})(\widehat{\mathcal{P}}_\perp^{(1)} \mathbf{u}_l^{(1)}, \ldots, \widehat{\mathcal{P}}_\perp^{(p)} \mathbf{u}_l^{(p)}, \widehat{\mathcal{P}}_\perp^{(1)} \mathbf{u}_l^{(1)}, \ldots, \widehat{\mathcal{P}}_\perp^{(p)} \mathbf{u}_l^{(p)}) \\
= & \sum_{1 \leq l' \leq r} \left[\sigma_{l'}^2 \prod_{q=1}^{p} \langle \widehat{\mathcal{P}}_\perp^{(q)} \mathbf{u}_l^{(q)}, \mathbf{u}_{l'}^{(q)} \rangle^2 \right] \\
\geq & \sigma_l^2 \prod_{q=1}^{p} \langle \widehat{\mathcal{P}}_\perp^{(q)} \mathbf{u}_l^{(q)}, \mathbf{u}_l^{(q)} \rangle^2 \\
= & \sigma_l^2 \prod_{q=1}^{p} \|\widehat{\mathcal{P}}_\perp^{(q)} \mathbf{u}_l^{(q)}\|^4.
\end{aligned}$$

根据归纳假设（17），

$$\|\widehat{\mathcal{P}}^{(q)} \mathbf{u}_l^{(q)}\|^2 = \sum_{1 \leq l' < k} \langle \mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_{l'}^{(q)} \rangle^2 \leq \sum_{1 \leq l' < k} \tilde{\delta}_{l'}^2 \leq C \delta_k^2.$$

通过将（ c_0 10）中的常数取得足够小，我们可以确保

$$\max_{l \notin \pi([k-1])} (\Sigma - \sigma_0^2 \mathscr{I})(\widehat{\mathcal{P}}_\perp^{(1)} \mathbf{u}_l^{(1)}, \ldots, \widehat{\mathcal{P}}_\perp^{(p)} \mathbf{u}_l^{(p)}, \widehat{\mathcal{P}}_\perp^{(1)} \mathbf{u}_l^{(1)}, \ldots, \widehat{\mathcal{P}}_\perp^{(p)} \mathbf{u}_l^{(p)}) \geq (1 - C\delta_k^4) \tau^2. \qquad (20)$$

其中

$$\tau^2 = \max_{l \notin \pi([k-1])} \sigma_l^2.$$

接下来，我们推导一个关于.....的界

$$\sup_{\|\mathbf{w}^{(q)}\| \leq 1, 1 \leq q \leq p} (\Sigma - \widehat{\Sigma})(\widehat{\mathcal{P}}_\perp^{(1)} \mathbf{w}^{(1)}, \ldots, \widehat{\mathcal{P}}_\perp^{(p)} \mathbf{w}^{(p)}, \widehat{\mathcal{P}}_\perp^{(1)} \mathbf{w}^{(1)}, \ldots, \widehat{\mathcal{P}}_\perp^{(p)} \mathbf{w}^{(p)}).$$

注意,

$$
\begin{aligned}
& (\Sigma-\widehat{\Sigma})(\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}^{(1)},\ldots,\widehat{\mathcal{P}}_{\perp}^{(p)}\mathbf{w}^{(p)},\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}^{(1)},\ldots,\widehat{\mathcal{P}}_{\perp}^{(p)}\mathbf{w}^{(p)}) \\
= & \sum_{l_1=1}^{r}\sum_{l_2=1}^{r}\sigma_{l_1}\sigma_{l_2}\left(\widehat{\Sigma}_{\theta,l_1l_2}-\Sigma_{\theta,l_1l_2}\right)\left(\prod_{q=1}^{p}\langle\mathbf{u}_{l_1}^{(q)},\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\rangle\right)\left(\prod_{q=1}^{p}\langle\mathbf{u}_{l_2}^{(q)},\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\rangle\right) \\
& +\frac{2}{n}\sum_{i=1}^{n}\sum_{l=1}^{r}\sigma_l\theta_{il}\left(\prod_{q=1}^{p}\langle\mathbf{u}_{l}^{(q)},\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\rangle\right)\mathscr{E}_i(\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}^{(1)},\ldots,\widehat{\mathcal{P}}_{\perp}^{(p)}\mathbf{w}^{(p)}) \\
& +\left(\frac{1}{n}\sum_{i=1}^{n}[\mathscr{E}_i(\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}^{(1)},\ldots,\widehat{\mathcal{P}}_{\perp}^{(p)}\mathbf{w}^{(p)})]^2-\sigma_0^2\prod_{q=1}^{p}\|\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\|^2\right). \qquad (21)
\end{aligned}
$$

我们分别对右侧的三项进行界定。

第一项可界定为

$$
\begin{aligned}
& \left|\sum_{l_1=1}^{r}\sum_{l_2=1}^{r}\sigma_{l_1}\sigma_{l_2}\left(\frac{1}{n}\sum_{i=1}^{n}\theta_{il_1}\theta_{il_2}\right)\left(\prod_{q=1}^{p}\langle\mathbf{u}_{l_1}^{(q)},\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\rangle\right)\left(\prod_{q=1}^{p}\langle\mathbf{u}_{l_2}^{(q)},\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\rangle\right)\right. \\
& \qquad \left.-\sum_{l=1}^{r}\sigma_l^2\left(\prod_{q=1}^{p}\langle\mathbf{u}_{l}^{(q)},\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\rangle\right)^2\right| \\
\leq & \|\widehat{\Sigma}_{\theta}-I_r\|\left[\sum_{l=1}^{r}\sigma_l^2\left(\prod_{q=1}^{p}\langle\mathbf{u}_{l}^{(q)},\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\rangle\right)^2\right] \\
= & \|\widehat{\Sigma}_{\theta}-I_r\|\left[\sum_{l\in\pi([k-1])}\sigma_l^2\left(\prod_{q=1}^{p}\langle\mathbf{u}_{l}^{(q)},\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\rangle\right)^2+\sum_{l\notin\pi([k-1])}\sigma_l^2\left(\prod_{q=1}^{p}\langle\mathbf{u}_{l}^{(q)},\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\rangle\right)^2\right],
\end{aligned}
$$

回忆一下，对于任意 $l\in\pi([k-1])$,

$$
|\langle\mathbf{u}_l^{(q)},\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\rangle|=|\langle\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{u}_l^{(q)},\mathbf{w}^{(q)}\rangle|\leq\|\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{u}_l^{(q)}\|\leq\delta_l.
$$

因此,

$$
\begin{aligned}
\sum_{l\in\pi([k-1])}\sigma_l^2\left(\prod_{q=1}^{p}\langle\mathbf{u}_{l}^{(q)},\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\rangle\right)^2 & \leq \max_{l\in\pi([k-1])}\sigma_l^2\left(\prod_{q=2}^{p}\langle\mathbf{u}_{l}^{(q)},\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\rangle\right)^2 \\
& \leq \max_{1\leq l<k}\{\sigma_{\pi(l)}^2\delta_l^{2p-2}\}\leq\max_{1\leq l<k}\{\sigma_{\pi(l)}^2\delta_l^2\}\leq\tau^2,
\end{aligned}
$$

通过取 c_0 使得 (10) 足够小。另一方面,

$$
\sum_{l\notin\pi([k-1])}\sigma_l^2\left(\prod_{q=1}^{p}\langle\mathbf{u}_{l}^{(q)},\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\rangle\right)^2\leq\max_{l\notin\pi([k-1])}\sigma_l^2=\tau^2.
$$

这意味着

$$\left|\sum_{l_1=1}^{r}\sum_{l_2=1}^{r}\sigma_{l_1}\sigma_{l_2}\left(\widehat{\Sigma}_{\theta,l_1l_2}-\Sigma_{\theta,l_1l_2}\right)\left(\prod_{q=1}^{p}\langle\mathbf{u}_{l_1}^{(q)},\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\rangle\right)\left(\prod_{q=1}^{p}\langle\mathbf{u}_{l_2}^{(q)},\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\rangle\right)\right|\leq C\tau^2\sqrt{\frac{r}{n}}. \tag{22}$$

同样，第二项可由下式界定

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{r}\sigma_l\theta_{il}\left(\prod_{q=1}^{p}\langle\mathbf{u}_{l}^{(q)},\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\rangle\right)\mathscr{E}_i(\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}^{(1)},\ldots,\widehat{\mathcal{P}}_{\perp}^{(p)}\mathbf{w}^{(p)})\right|\\ \leq\ &\|\widehat{\Sigma}_{\theta,\mathscr{E}}\|\left[\sum_{l=1}^{r}\sigma_l^2\left(\prod_{q=1}^{p}\langle\mathbf{u}_{l}^{(q)},\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\rangle\right)^2\right]^{1/2}\\ \leq\ &C\tau\sigma_0\sqrt{\frac{d}{n}}.\end{aligned} \tag{23}$$

最后，第三项可由下式界定

$$\left|\frac{1}{n}\sum_{i=1}^{n}[\mathscr{E}_i(\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}^{(1)},\ldots,\widehat{\mathcal{P}}_{\perp}^{(p)}\mathbf{w}^{(p)})]^2-\sigma_0^2\prod_{q=1}^{p}\|\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{w}^{(q)}\|^2\right|\leq\|\widehat{\Sigma}_{\mathscr{E}}-\Sigma_{\mathscr{E}}\|\leq C\sigma_0^2\sqrt{\frac{d}{n}}. \tag{24}$$

结合（21）—（24），可得

$$\begin{aligned}&\sup_{\|\mathbf{w}^{(q)}\|\leq1,1\leq q\leq p}\left|(\widehat{\Sigma}-\Sigma)(\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}^{(1)},\ldots,\widehat{\mathcal{P}}_{\perp}^{(p)}\mathbf{w}^{(p)},\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}^{(1)},\ldots,\widehat{\mathcal{P}}_{\perp}^{(p)}\mathbf{w}^{(p)})\right|\\ \leq\ &C\tau^2\sqrt{\frac{r}{n}}+C(\sigma_0\tau+\sigma_0^2)\sqrt{\frac{d}{n}}.\end{aligned}$$

结合（19），可推出

$$\tilde{\Sigma}(\hat{\mathbf{u}}_k^{(1)},\hat{\mathbf{u}}_k^{(2)},\ldots,\hat{\mathbf{u}}_k^{(p)},\hat{\mathbf{u}}_k^{(1)},\ldots,\hat{\mathbf{u}}_k^{(p)})\geq\tau^2\left(1-C\delta_k^4-C\sqrt{\frac{r}{n}}\right)-C(\sigma_0\tau+\sigma_0^2)\sqrt{\frac{d}{n}}-C\sigma_0^2\delta_k^2, \tag{25}$$

通过取 c_0 使（10）足够小。

B.2 $\tilde{\Sigma}\left(\widehat{\mathcal{P}}_{\perp}^{(1)}\left(\hat{\mathbf{u}}_k^{(1)}-\langle\hat{\mathbf{u}}_k^{(1)},\mathbf{u}_{\pi(k)}^{(1)}()\rangle\mathbf{u}_{\pi(k)}^{(1)}()\right)\hat{\mathbf{u}}_k^{(2)},\ldots,\hat{\mathbf{u}}_k^{(p)},\hat{\mathbf{u}}_k^{(1)},\ldots,\hat{\mathbf{u}}_k^{(p)}\right)$

注意，对于任意与$\mathbf{w}$ 正交的$\mathbf{u}_{\pi(k)}^{(1)}()$，我们有

$$
\begin{aligned}
&\tilde{\Sigma}(\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w},\hat{\mathbf{u}}_k^{(2)},\ldots,\hat{\mathbf{u}}_k^{(p)},\hat{\mathbf{u}}_k^{(1)},\ldots,\hat{\mathbf{u}}_k^{(p)})\\
=\;&\sigma_{\pi(k)}^2\left(\frac{1}{n}\sum_{i=1}^n\theta_{i\pi(k)}^2\right)\left(\prod_{q=2}^p\langle\mathbf{u}_{\pi(k)}^{(q)},\hat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^p\langle\mathbf{u}_{\pi(k)}^{(q)},\hat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_{\pi(k)}^{(1)},\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}\rangle\\
&+\sum_{l\neq\pi(k)}\sigma_{\pi(k)}\sigma_l\left(\frac{1}{n}\sum_{i=1}^n\theta_{i\pi(k)}\theta_{il}\right)\left(\prod_{q=2}^p\langle\mathbf{u}_{\pi(k)}^{(q)},\hat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^p\langle\mathbf{u}_l^{(q)},\hat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_{\pi(k)}^{(1)},\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}\rangle\\
&+\sum_{l\neq\pi(k)}\sigma_l\sigma_{\pi(k)}\left(\frac{1}{n}\sum_{i=1}^n\theta_{il}\theta_{i\pi(k)}\right)\left(\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\hat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^p\langle\mathbf{u}_{\pi(k)}^{(q)},\hat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}\rangle\\
&+\sum_{l_1,l_2\neq\pi(k)}\sigma_{l_1}\sigma_{l_2}\left(\frac{1}{n}\sum_{i=1}^n\theta_{il_1}\theta_{il_2}\right)\left(\prod_{q=2}^p\langle\mathbf{u}_{l_1}^{(q)},\hat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^p\langle\mathbf{u}_{l_2}^{(q)},\hat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_{l_1}^{(1)},\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}\rangle\\
&+\frac{1}{n}\sum_{i=1}^n\left[\sum_{l=1}^r\sigma_l\theta_{il}\left(\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\hat{\mathbf{u}}_k^{(q)}\rangle\right)\mathscr{E}_i(\hat{\mathbf{u}}_k^{(1)},\ldots,\hat{\mathbf{u}}_k^{(p)})\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}\rangle\right]\\
&+\frac{1}{n}\sum_{i=1}^n\left[\sum_{l=1}^r\mathscr{E}_i(\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w},\hat{\mathbf{u}}_k^{(2)},\ldots,\hat{\mathbf{u}}_k^{(p)})\sigma_l\theta_{il}\left(\prod_{q=1}^p\langle\mathbf{u}_l^{(q)},\hat{\mathbf{u}}_k^{(q)}\rangle\right)\right]\\
&+\frac{1}{n}\sum_{i=1}^n\mathscr{E}_i(\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w},\hat{\mathbf{u}}_k^{(2)},\ldots,\hat{\mathbf{u}}_k^{(p)})\mathscr{E}_i(\hat{\mathbf{u}}_k^{(1)},\ldots,\hat{\mathbf{u}}_k^{(p)})-\sigma_0^2\langle\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w},\hat{\mathbf{u}}_k^{(p)}\rangle. \qquad (26)
\end{aligned}
$$

同样，右侧的每一项都需要仔细进行界定。

右侧的第一项（26）可由下式界定

$$
\begin{aligned}
&\left|\sigma_{\pi(k)}^2\left(\frac{1}{n}\sum_{i=1}^n\theta_{i\pi(k)}^2\right)\left(\prod_{q=2}^p\langle\mathbf{u}_{\pi(k)}^{(q)},\hat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^p\langle\mathbf{u}_{\pi(k)}^{(q)},\hat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_{\pi(k)}^{(1)},\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}\rangle\right|\\
&\qquad\qquad\leq\tau^2\|\widehat{\Sigma}_\theta\|_{\max}|\langle\mathbf{u}_{\pi(k)}^{(1)},\widehat{\mathcal{P}}_{\perp}\mathbf{w}\rangle|.
\end{aligned}
$$

特别地，当$\mathbf{w}=\hat{\mathbf{u}}_k^{(1)}-\langle\hat{\mathbf{u}}_k^{(1)},\mathbf{u}_{\pi(k)}^{(1)}()\rangle\mathbf{u}_{\pi(k)}^{(1)}()$时，我们有

$$
\begin{aligned}
|\langle\mathbf{u}_{\pi(k)}^{(1)},\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}\rangle| &= \left|\langle\hat{\mathbf{u}}_k^{(1)},\mathbf{u}_{\pi(k)}^{(1)}\rangle(1-\langle\mathbf{u}_{\pi(k)}^{(1)},\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{u}_{\pi(k)}^{(1)}\rangle)\right|\\
&\leq 1-\langle\mathbf{u}_{\pi(k)}^{(1)},\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{u}_{\pi(k)}^{(1)}\rangle\\
&= \|\widehat{\mathcal{P}}^{(1)}\mathbf{u}_{\pi(k)}^{(1)}\|^2\leq\sum_{1\leq l<k}\tilde{\delta}_l^2\leq C\delta_k^2,
\end{aligned}
$$

通过取足够小的 c_0 。因此，

$$
\left|\sigma_{\pi(k)}^2\left(\frac{1}{n}\sum_{i=1}^n\theta_{i\pi(k)}^2\right)\left(\prod_{q=2}^p\langle\mathbf{u}_{\pi(k)}^{(q)},\hat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^p\langle\mathbf{u}_{\pi(k)}^{(q)},\hat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_{\pi(k)}^{(1)},\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}\rangle\right|\leq C\tau^2\delta_k^2. \qquad (27)
$$

右侧的第二项（26）可按如下方式进行界定：

$$
\begin{aligned}
&\left|\sum_{l\neq\pi(k)}\sigma_{\pi(k)}\sigma_l\left(\frac{1}{n}\sum_{i=1}^{n}\theta_{i\pi(k)}\theta_{il}\right)\left(\prod_{q=2}^{p}\langle\mathbf{u}_{\pi(k)}^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^{p}\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_{\pi(k)}^{(1)},\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}\rangle\right| \\
\leq\ & \tau\|\widehat{\Sigma}_\theta-I\|\left(\sum_{l\neq\pi(k)}\sigma_l^2\prod_{q=1}^{p}\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle^2\right)^{1/2} \\
\leq\ & \tau\|\widehat{\Sigma}_\theta-I\|\left(\sum_{l\neq\pi(k)}\langle\mathbf{u}_l^{(1)},\widehat{\mathbf{u}}_k^{(1)}\rangle^2\right)^{1/2}\max_{l\neq\pi(k)}\left\{\sigma_l\prod_{q=2}^{p}|\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle|\right\} \\
=\ & \tau\delta_k\|\widehat{\Sigma}_\theta-I\|\max_{l\neq\pi(k)}\left\{\sigma_l\prod_{q=2}^{p}|\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle|\right\} \\
\leq\ & \tau\delta_k\|\widehat{\Sigma}_\theta-I\|\max\left\{\max_{l\in\pi([k-1])}\left\{\sigma_l\prod_{q=2}^{p}|\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle|\right\},\max_{l\notin\pi([k])}\left\{\sigma_l\prod_{q=2}^{p}|\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle|\right\}\right\}.
\end{aligned}
$$

注意，

$$
\max_{l\in\pi([k-1])}\left\{\sigma_l\prod_{q=2}^{p}|\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle|\right\}\leq\max_{l\in\pi([k-1])}\left\{\sigma_l\prod_{q=2}^{p}|\langle\widehat{\mathcal{P}}_{\perp}^{(q)}\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle|\right\}\leq\max_{1\leq l<k}\{\sigma_{\pi(l)}\delta_l^{p-1}\}\leq 2\tau\delta_k^{p-1},
$$

and

$$
\max_{l\notin\pi([k])}\left\{\sigma_l\prod_{q=2}^{p}|\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle|\right\}\leq\tau\delta_k^{p-1}.
$$

我们得到

$$
\left|\sum_{l\neq\pi(k)}\sigma_{\pi(k)}\sigma_l\left(\frac{1}{n}\sum_{i=1}^{n}\theta_{i\pi(k)}\theta_{il}\right)\left(\prod_{q=2}^{p}\langle\mathbf{u}_{\pi(k)}^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^{p}\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\right|\leq C\tau^2\delta_k^p\sqrt{\frac{r}{n}}. \tag{28}
$$

同样地，当 $\mathbf{w}=\widehat{}\mathbf{u}_k^{(1)}-\langle\widehat{}\mathbf{u}_k^{(1)}$，$\mathbf{u}_{\pi\ k()}^{(1)}\rangle\mathbf{u}_{\pi\ k()}^{(1)}$时，我们将（26）右侧的第三项限定为

$$
\begin{aligned}
&\left|\sum_{l\neq\pi(k)}\sigma_l\sigma_{\pi(k)}\left(\frac{1}{n}\sum_{i=1}^{n}\theta_{il}\theta_{i\pi(k)}\right)\left(\prod_{q=2}^{p}\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^{p}\langle\mathbf{u}_{\pi(k)}^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}\rangle\right| \\
\leq\ & \tau\|\widehat{\Sigma}_\theta-I\|\left(\sum_{l\neq\pi(k)}\sigma_l^2\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}\rangle^2\prod_{q=2}^{p}\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle^2\right)^{1/2} \\
\leq\ & \tau\|\widehat{\Sigma}_\theta-I\|\left(\sum_{l\neq\pi(k)}\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{w}\rangle^2\right)^{1/2}\max_{l\neq\pi(k)}\left\{\sigma_l\prod_{q=2}^{p}|\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right\} \\
\leq\ & C\tau^2\delta_k^p\sqrt{\frac{r}{n}},
\end{aligned} \tag{29}
$$

其中在最后一个不等式中，我们使用了以下事实

$$\sum_{l\neq\pi(k)}\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w}\rangle^2\leq\|\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w}\|^2\leq\|\mathbf{w}\|^2\leq\delta_k^2;$$

将第四项替换为

$$\begin{aligned}
&\left|\sum_{l_1,l_2\neq\pi(k)}\sigma_{l_1}\sigma_{l_2}\left(\frac{1}{n}\sum_{i=1}^n\theta_{il_1}\theta_{il_2}\right)\left(\prod_{q=2}^p\langle\mathbf{u}_{l_1}^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^p\langle\mathbf{u}_{l_2}^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_{l_1}^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w}\rangle\right| \\
\leq\ & \|\widehat{\Sigma}_\theta\|\left(\sum_{l\neq\pi(k)}\sigma_l^2\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w}\rangle^2\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle^2\right)^{1/2}\left(\sum_{l\neq\pi(k)}\sigma_l^2\prod_{q=1}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle^2\right)^{1/2} \\
\leq\ & C\tau^2\delta_k^{2(p-1)};
\end{aligned}\tag{30}$$

将第五项替换为

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n\left[\sum_{l\neq\pi(k)}\sigma_l\theta_{il}\left(\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\mathscr{E}_i(\widehat{\mathbf{u}}_k^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w}\rangle\right] \\
=\ & \sum_{l\neq\pi(k)}\left\{\sigma_l\left(\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w}\rangle\left[\frac{1}{n}\sum_{i=1}^n\theta_{il}\mathscr{E}_i(\widehat{\mathbf{u}}_k^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\right]\right\} \\
\leq\ & \|\widehat{\Sigma}_{\theta,\mathscr{E}}\|\cdot\left[\sum_{l\neq\pi(k)}\left(\sigma_l^2\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w}\rangle^2\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle^2\right)\right]^{1/2} \\
\leq\ & C\sigma_0\tau\delta_k^{p-1}\sqrt{\frac{d}{n}};
\end{aligned}\tag{31}$$

并将第六项替换为

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n\left[\sum_{l=1}^r\mathscr{E}_i(\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\sigma_l\theta_{il}\left(\prod_{q=1}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\right] \\
\leq\ & \frac{1}{n}\sum_{i=1}^n\left[\mathscr{E}_i(\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\sigma_{\pi(k)}\theta_{i\pi(k)}\left(\prod_{q=1}^p\langle\mathbf{u}_{\pi(k)}^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\right] \\
&+\frac{1}{n}\sum_{i=1}^n\left[\sum_{l\neq\pi(k)}\mathscr{E}_i(\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\sigma_l\theta_{il}\left(\prod_{q=1}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\right] \\
\leq\ & \tau\|\widehat{\Sigma}_{\theta,\mathscr{E}}\|+\|\widehat{\Sigma}_{\theta,\mathscr{E}}\|\left(\sum_{l\neq\pi(k)}\sigma_l^2\prod_{q=1}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle^2\right)^{1/2} \\
\leq\ & C\sigma_0\tau\sqrt{\frac{d}{n}}.
\end{aligned}\tag{32}$$

最后一项可以由以下方式界定

$$\left|\frac{1}{n}\sum_{i=1}^n\mathscr{E}_i(\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\mathscr{E}_i(\widehat{\mathbf{u}}_k^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(p)})-\sigma_0^2\langle\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w},\widehat{\mathbf{u}}_k^{(1)}\rangle\right|\leq\|\widehat{\Sigma}_{\mathscr{E}}-\Sigma_{\mathscr{E}}\|\leq C\sigma_0^2\sqrt{\frac{d}{n}}.$$

结合（27）-（32），我们得到

$$\tilde{\Sigma}(\widehat{\mathcal{P}}_{\perp}^{(1)}(\widehat{\mathbf{u}}_k^{(1)} - \langle\widehat{\mathbf{u}}_k^{(1)}, \mathbf{u}_{\pi(k)}^{(1)}\rangle\mathbf{u}_{\pi(k)}^{(1)}), \widehat{\mathbf{u}}_k^{(2)}, \ldots, \widehat{\mathbf{u}}_k^{(p)}, \widehat{\mathbf{u}}_k^{(1)}, \ldots, \widehat{\mathbf{u}}_k^{(p)}) \leq C\tau^2\delta_k^2 + C(\sigma_0^2 + \sigma_0\tau)\sqrt{\frac{d}{n}}.$$

B.3 上界 最大值$_{l\notin\pi([k])}\tilde{\Sigma}(\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{u}_m^{(1)}, \widehat{\mathbf{u}}_k^{(2)}, \ldots, \widehat{\mathbf{u}}_k^{(p)}, \widehat{\mathbf{u}}_k^{(1)}, \ldots, \widehat{\mathbf{u}}_k^{(p)})$

为了推导辅助界（17），我们还需要一个上界来约束

$$\max_{l\notin\pi([k])} \tilde{\Sigma}(\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{u}_m^{(1)}, \widehat{\mathbf{u}}_k^{(2)}, \ldots, \widehat{\mathbf{u}}_k^{(p)}, \widehat{\mathbf{u}}_k^{(1)}, \ldots, \widehat{\mathbf{u}}_k^{(p)}).$$

我们将采用类似的步骤，对（26）右侧的每一项分别给出界，但现在使用 $\mathbf{w} = \mathbf{u}_m^{(1)}$($m \notin \pi([k])$).

具体而言，第一项可以由下式给出上界

$$\left|\sigma_{\pi(k)}^2\left(\frac{1}{n}\sum_{i=1}^n\theta_{i\pi(k)}^2\right)\left(\prod_{q=2}^p\langle\mathbf{u}_{\pi(k)}^{(q)}, \widehat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^p\langle\mathbf{u}_{\pi(k)}^{(q)}, \widehat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_{\pi(k)}^{(1)}, \widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{u}_m^{(1)}\rangle\right| \leq \tau^2\|\widehat{\Sigma}_\theta\|_{\max}|\langle\mathbf{u}_{\pi(k)}^{(1)}, \widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{u}_m^{(1)}\rangle|.$$

注意

$$\langle\mathbf{u}_{\pi(k)}^{(1)}, \mathbf{u}_m^{(1)}\rangle = \langle\mathbf{u}_{\pi(k)}^{(1)}, \widehat{\mathcal{P}}^{(1)}\mathbf{u}_m^{(1)}\rangle + \langle\mathbf{u}_{\pi(k)}^{(1)}, \widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{u}_m^{(1)}\rangle = 0.$$

我们得到

$$|\langle\mathbf{u}_{\pi(k)}^{(1)}, \widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{u}_m^{(1)}\rangle| = |\langle\mathbf{u}_{\pi(k)}^{(1)}, \widehat{\mathcal{P}}^{(1)}\mathbf{u}_m^{(1)}\rangle| \leq \|\widehat{\mathcal{P}}\mathbf{u}_{\pi(k)}^{(1)}\|\|\widehat{\mathcal{P}}\mathbf{u}_m^{(1)}\| \leq \sum_{1\leq l<k}\tilde{\delta}_l^2 \leq C\delta_k^2,$$

由柯西-施瓦茨不等式可得。这意味着

$$\left|\sigma_{\pi(k)}^2\left(\frac{1}{n}\sum_{i=1}^n\theta_{i\pi(k)}^2\right)\left(\prod_{q=2}^p\langle\mathbf{u}_{\pi(k)}^{(q)}, \widehat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^p\langle\mathbf{u}_{\pi(k)}^{(q)}, \widehat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_{\pi(k)}^{(1)}, \widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{u}_m^{(1)}\rangle\right| \leq C\tau^2\delta_k^2.$$

只需将 c_0 取足够小即可。

第二项也可以通过以下方式界定:

$$\begin{aligned}
&\left|\sum_{l\neq\pi(k)}\sigma_{\pi(k)}\sigma_l\left(\frac{1}{n}\sum_{i=1}^n\theta_{i\pi(k)}\theta_{il}\right)\left(\prod_{q=2}^p\langle\mathbf{u}_{\pi(k)}^{(q)}, \widehat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^p\langle\mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_{\pi(k)}^{(1)}, \widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{u}_m^{(1)}\rangle\right| \\
\leq\ & \tau\|\widehat{\Sigma}_\theta - I\||\langle\mathbf{u}_{\pi(k)}^{(1)}, \widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{u}_m^{(1)}\rangle|\left(\sum_{l\neq\pi(k)}\sigma_l^2\prod_{q=1}^p\langle\mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_k^{(q)}\rangle^2\right)^{1/2} \\
\leq\ & \tau^2\delta_k^{p+1}\sqrt{\frac{r}{n}}.
\end{aligned}$$

我们通过以下方式界定第三项:

$$\begin{aligned}
&\left|\sum_{l\neq\pi(k)}\sigma_l\sigma_{\pi(k)}\left(\frac{1}{n}\sum_{i=1}^n\theta_{il}\theta_{i\pi(k)}\right)\left(\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^p\langle\mathbf{u}_{\pi(k)}^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_m^{(1)}\rangle\right| \\
\leq\ &\left|\sigma_m\sigma_{\pi(k)}\left(\frac{1}{n}\sum_{i=1}^n\theta_{im}\theta_{i\pi(k)}\right)\left(\prod_{q=2}^p\langle\mathbf{u}_m^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^p\langle\mathbf{u}_{\pi(k)}^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_m^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_m^{(1)}\rangle\right| \\
&+\left|\sum_{l\notin\{\pi(k),m\}}\sigma_l\sigma_{\pi(k)}\left(\frac{1}{n}\sum_{i=1}^n\theta_{il}\theta_{i\pi(k)}\right)\left(\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^p\langle\mathbf{u}_{\pi(k)}^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_m^{(1)}\rangle\right|
\end{aligned}$$

右侧的第一项还可以进一步界定为

$$\tau^2\|\widehat{\Sigma}_\theta-\Sigma_\theta\|_{\max}\prod_{q=2}^p\left|\langle\mathbf{u}_m^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right|\leq C\tau^2\delta_k^{p-1}\sqrt{\frac{\log r}{n}}.$$

现在考虑第二项:

$$\begin{aligned}
&\left|\sum_{l\notin\{\pi(k),m\}}\sigma_l\sigma_{\pi(k)}\left(\frac{1}{n}\sum_{i=1}^n\theta_{il}\theta_{i\pi(k)}\right)\left(\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^p\langle\mathbf{u}_{\pi(k)}^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_m^{(1)}\rangle\right| \\
\leq\ &\tau\|\widehat{\Sigma}_\theta-I\|\left(\sum_{l\notin\{\pi(k),m\}}\sigma_l^2\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_m^{(1)}\rangle^2\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle^2\right)^{1/2} \\
\leq\ &\tau\|\widehat{\Sigma}_\theta-I\|\left(\sum_{l\in\pi([k-1])}\sigma_l^2\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_m^{(1)}\rangle^2\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle^2\right. \\
&\qquad\left.+\sum_{l\notin\pi([k])\cup\{m\}}\sigma_l^2\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_m^{(1)}\rangle^2\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle^2\right)^{1/2}.
\end{aligned}$$

最右侧括号中的第一项可以界定为

$$\begin{aligned}
&\sum_{l\in\pi([k-1])}\sigma_l^2\|\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_l^{(1)}\|^2\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle^2 \\
\leq\ &\left(\sum_{l\in\pi([k-1])}\langle\mathbf{u}_l^{(2)},\widehat{\mathbf{u}}_k^{(2)}\rangle^2\right)\left(\max_{l\in\pi([k-1])}\sigma_l^2\|\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_l^{(1)}\|^2\prod_{q=3}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle^2\right) \\
\leq\ &\delta_k^2\left(\max_{l\in\pi([k-1])}\sigma_l^2\delta_l^2\delta_k^{2(p-2)}\right)\leq C\tau^2\delta_k^{2p};
\end{aligned}$$

第二项乘以

$$\|\widehat{\mathcal{P}}^{(1)}\mathbf{u}_m^{(1)}\|^2\left(\max_{l\notin\pi([k])\cup\{m\}}\sigma_l^2\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle^2\right)\leq C\tau^2\delta_k^{2p},$$

以便

$$\begin{aligned}
&\left|\sum_{l\neq\pi(k)}\sigma_l\sigma_{\pi(k)}\left(\frac{1}{n}\sum_{i=1}^n\theta_{il}\theta_{i\pi(k)}\right)\left(\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^p\langle\mathbf{u}_{\pi(k)}^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_m^{(1)}\rangle\right| \\
\leq\ & C\left(\tau^2\delta_k^{p-1}\sqrt{\frac{\log r}{n}}+\tau^2\delta_k^p\sqrt{\frac{r}{n}}\right).
\end{aligned}$$

与.....类似 在此之前，等式右侧的第四项（26）可以由以下内容界定：

$$\begin{aligned}
&\left|\sum_{l_1,l_2\neq\pi(k)}\sigma_{l_1}\sigma_{l_2}\left(\frac{1}{n}\sum_{i=1}^n\theta_{il_1}\theta_{il_2}\right)\left(\prod_{q=2}^p\langle\mathbf{u}_{l_1}^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\left(\prod_{q=1}^p\langle\mathbf{u}_{l_2}^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_{l_1}^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_m^{(1)}\rangle\right| \\
\leq\ & \|\widehat{\Sigma}_\theta\|\left(\sum_{l\neq\pi(k)}\sigma_l^2\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_m^{(1)}\rangle^2\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle^2\right)^{1/2}\left(\sum_{l\neq\pi(k)}\sigma_l^2\prod_{q=1}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle^2\right)^{1/2} \\
\leq\ & C\tau^2\delta_k^{2(p-1)};
\end{aligned}$$

而第五项可以由以下内容界定：

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^n\left[\sum_{l\neq\pi(k)}\sigma_l\theta_{il}\left(\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\mathscr{E}_i(\widehat{\mathbf{u}}_k^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_m^{(1)}\rangle\right]\right| \\
=\ & \left|\sum_{l\neq\pi(k)}\left\{\sigma_l\left(\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_m^{(1)}\rangle\left[\frac{1}{n}\sum_{i=1}^n\theta_{il}\mathscr{E}_i(\widehat{\mathbf{u}}_k^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\right]\right\}\right| \\
\leq\ & \|\widehat{\Sigma}_{\theta,\mathscr{E}}\|\cdot\left(\sum_{l\neq\pi(k)}\sigma_l^2\langle\mathbf{u}_l^{(1)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_m^{(1)}\rangle^2\prod_{q=2}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle^2\right)^{1/2} \\
\leq\ & C\sigma_0\tau\delta_k^{p-1}\sqrt{\frac{d}{n}}.
\end{aligned}$$

现在我们转向（26）右侧的第六项。记

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n\left[\sum_{l=1}^r\mathscr{E}_i(\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\sigma_l\theta_{il}\left(\prod_{q=1}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\right] \\
\leq\ & \frac{1}{n}\sum_{i=1}^n\left[\sum_{l\neq\pi(k)}\mathscr{E}_i(\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\sigma_l\theta_{il}\left(\prod_{q=1}^p\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\right] \\
&+\frac{1}{n}\sum_{i=1}^n\left[\mathscr{E}_i(\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\sigma_{\pi(k)}\theta_{i\pi(k)}\left(\prod_{q=1}^p\langle\mathbf{u}_{\pi(k)}^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\right].
\end{aligned}$$

第一项同样可以由以下式子界定

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}\left[\sum_{l\neq\pi(k)}\mathscr{E}_i(\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\sigma_l\theta_{il}\left(\prod_{q=1}^{p}\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\right]\right| \\
\leq\ & \|\widehat{\Sigma}_{\theta,\mathscr{E}}\|\left(\sum_{l\neq\pi(k)}\sigma_l^2\prod_{q=1}^{p}\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle^2\right)^{1/2}\leq C\sigma_0\tau\delta_k^{p-1}\sqrt{\frac{d}{n}}.
\end{aligned}$$

对于第二项，注意到

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}\left[\mathscr{E}_i(\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\sigma_{\pi(k)}\theta_{i\pi(k)}\left(\prod_{q=1}^{p}\langle\mathbf{u}_{\pi(k)}^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\right]\right| \\
\leq\ & \tau\left|\frac{1}{n}\sum_{i=1}^{n}\left[\mathscr{E}_i(\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\theta_{i\pi(k)}\right]\right| \\
\leq\ & \tau\left|\frac{1}{n}\sum_{i=1}^{n}\left[\mathscr{E}_i(\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\theta_{i\pi(k)}\right]\right|+\tau\left|\frac{1}{n}\sum_{i=1}^{n}\left[\mathscr{E}_i(\widehat{\mathcal{P}}^{(1)}\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\theta_{i\pi(k)}\right]\right| \quad (33)
\end{aligned}$$

其中第二个不平等由三角不等式得出。与前面一样,

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left[\mathscr{E}_i(\widehat{\mathcal{P}}^{(1)}\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\theta_{i\pi(k)}\right]\right|\leq\|\widehat{\Sigma}_{\theta,\mathscr{E}}\|\|\widehat{\mathcal{P}}^{(1)}\mathbf{u}_m^{(1)}\|\leq C\sigma_0\delta_k\sqrt{\frac{d}{n}}.$$

为了对最右端的第一项进行界定（33），令

$$\widehat{\mathbf{u}}_k^{(q)}=\alpha_q\mathbf{u}_{\pi(k)}+\mathbf{v}^{(q)}$$

其中 $\alpha_q=\langle\mathbf{u}^{\pi(k)},\widehat{\ }\mathbf{u}_k\notin\{\pi(k)\ m\}\notin\{\pi(k)\ m\}\notin\{\pi(k)\ m\}^{(q)}\rangle$ 且 $\|\mathbf{v}^{(q)}\|=\sqrt{1-\alpha_q^2}\leq\delta_{k\circ}$ 那么

$$\begin{aligned}
&\mathscr{E}_i(\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)}) \\
=\ & \alpha_p\mathscr{E}_i(\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p-1)},\mathbf{u}_{\pi(k)}^{(p)})+\mathscr{E}_i(\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p-1)},\mathbf{v}^{(p)}) \\
=\ & \alpha_p\alpha_{p-1}\mathscr{E}_i(\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p-2)},\mathbf{u}_{\pi(k)}^{(p-1)},\mathbf{u}_{\pi(k)}^{(p)}) \\
&+\alpha_p\mathscr{E}_i(\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p-2)},\mathbf{v}^{(p-1)},\mathbf{u}_{\pi(k)}^{(p)}) \\
&+\mathscr{E}_i(\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p-1)},\mathbf{v}^{(p)}) \\
=\ & \cdots\cdots \\
=\ & \left(\prod_{q=2}^{p}\alpha_q\right)\mathscr{E}_i(\mathbf{u}_m^{(1)},\mathbf{u}_{\pi(k)}^{(2)},\ldots,\mathbf{u}_{\pi(k)}^{(p)}) \\
&+\cdots+\mathscr{E}_i(\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p-1)},\mathbf{v}^{(p)}).
\end{aligned}$$

因此，

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}\left[\mathscr{E}_i(\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\theta_{i\pi(k)}\right]\right| \\
\leq\ &\left|\frac{1}{n}\sum_{i=1}^{n}\left[\mathscr{E}_i(\mathbf{u}_m^{(1)},\mathbf{u}_{\pi(k)}^{(2)},\ldots,\mathbf{u}_{\pi(k)}^{(p)})\theta_{i\pi(k)}\right]\right|+\cdots+\left|\frac{1}{n}\sum_{i=1}^{n}\left[\mathscr{E}_i(\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p-1)},\mathbf{v}^{(p)})\theta_{i\pi(k)}\right]\right| \\
\leq\ &\left|\frac{1}{n}\sum_{i=1}^{n}\left[\mathscr{E}_i(\mathbf{u}_m^{(1)},\mathbf{u}_{\pi(k)}^{(2)},\ldots,\mathbf{u}_{\pi(k)}^{(p)})\theta_{i\pi(k)}\right]\right|+C_p\delta_k\|\widehat{\Sigma}_{\theta,\mathscr{E}}\| \\
\leq\ &C\left(\sigma_0\sqrt{\frac{\log r}{n}}+\sigma_0\delta_k\sqrt{\frac{d}{n}}\right),
\end{aligned}$$

从而，26）最右端的第六项可由以下表达式上界：

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left[\sum_{l=1}^{r}\mathscr{E}_i(\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\sigma_l\theta_{il}\left(\prod_{q=1}^{p}\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right)\right]\right|\leq C\left(\sigma_0\tau\sqrt{\frac{\log r}{n}}+\sigma_0\tau\delta_k\sqrt{\frac{d}{n}}\right).$$

最后考虑第七项：

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}\mathscr{E}_i(\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\mathscr{E}_i(\widehat{\mathbf{u}}_k^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(p)})-\langle\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(1)}\rangle\right| \\
\leq\ &\left|\frac{1}{n}\sum_{i=1}^{n}\mathscr{E}_i(\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\mathscr{E}_i(\widehat{\mathbf{u}}_k^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(p)})-\langle\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(1)}\rangle\right| \\
&+\left|\frac{1}{n}\sum_{i=1}^{n}\mathscr{E}_i(\widehat{\mathcal{P}}^{(1)}\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\mathscr{E}_i(\widehat{\mathbf{u}}_k^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(p)})-\langle\widehat{\mathcal{P}}^{(1)}\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(1)}\rangle\right|.
\end{aligned}$$

与之前类似，第二项可以通过以下方式进行界定

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}\mathscr{E}_i(\widehat{\mathcal{P}}^{(1)}\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\mathscr{E}_i(\widehat{\mathbf{u}}_k^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(p)})-\sigma_0^2\langle\widehat{\mathcal{P}}^{(1)}\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(1)}\rangle\right| \\
\leq\ &\|\widehat{\mathcal{P}}^{(1)}\mathbf{u}_m^{(1)}\|\|\widehat{\Sigma}_{\mathscr{E}}-\Sigma_{\mathscr{E}}\|\leq C\sigma_0^2\delta_k\sqrt{\frac{d}{n}};
\end{aligned}$$

第一项可以通过以下方式界定

$$
\begin{aligned}
& \left|\frac{1}{n}\sum_{i=1}^{n}\mathscr{E}_i(\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\mathscr{E}_i(\widehat{\mathbf{u}}_k^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(p)})-\sigma_0^2\langle\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(1)}\rangle\right| \\
\leq\ & \left|\frac{1}{n}\sum_{i=1}^{n}\mathscr{E}_i(\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\mathscr{E}_i(\mathbf{v}^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(p)})-\sigma_0^2\langle\mathbf{u}_m^{(1)},\mathbf{v}^{(1)}\rangle\right| \\
& +\left|\frac{1}{n}\sum_{i=1}^{n}\mathscr{E}_i(\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\mathscr{E}_i(\mathbf{u}_{\pi(k)}^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\right| \\
\leq\ & C\delta_k\|\widehat{\Sigma}_{\mathscr{E}}-\Sigma_{\mathscr{E}}\|+\left|\frac{1}{n}\sum_{i=1}^{n}\mathscr{E}_i(\mathbf{u}_m^{(1)},\mathbf{u}_{\pi(k)}^{(2)},\cdots,\mathbf{u}_{\pi(k)}^{(p)})\mathscr{E}_i(\mathbf{u}_{\pi(k)}^{(1)},\cdots,\mathbf{u}_{\pi(k)}^{(p)})\right| \\
& +\cdots+\left|\frac{1}{n}\sum_{i=1}^{n}\mathscr{E}_i(\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\mathscr{E}_i(\mathbf{u}_{\pi(k)}^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p-1)},\mathbf{v}^{(p)})\right| \\
\leq\ & C\left(\sigma_0^2\delta_k\sqrt{\frac{d}{n}}+\sigma_0^2\sqrt{\frac{\log r}{n}}\right).
\end{aligned}
$$

将所有七个上界合并起来，我们得到

$$
\max_{l\notin\pi([k])}\tilde{\Sigma}(\widehat{\mathcal{P}}_{\perp}^{(1)}\mathbf{u}_m^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)},\widehat{\mathbf{u}}_k^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\leq C\left(\tau^2\delta_k^2+(\tau^2\delta_k^{p-1}+\sigma_0\tau+\sigma_0^2)\sqrt{\frac{\log r}{n}}\right).
$$

B.4 收尾

我们首先验证（18）。注意

$$
\begin{aligned}
& (\Sigma-\Sigma_{\mathscr{E}})(\widehat{\mathbf{u}}_k^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)},\widehat{\mathbf{u}}_k^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(p)}) \\
=\ & \sum_{l=1}^{r}\left(\sigma_l^2\prod_{q=1}^{p}\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle^2\right) \\
\leq\ & \max_{1\leq l\leq r}\left\{\sigma_l^2\left|\prod_{q=1}^{p}\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right|\right\}\cdot\left(\sum_{l=1}^{r}\left|\prod_{q=1}^{p}\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right|\right) \\
\leq\ & \max_{1\leq l\leq r}\left\{\sigma_l^2\left|\prod_{q=1}^{p}\langle\mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_k^{(q)}\rangle\right|\right\}.
\end{aligned}
$$

其中最后一个在 等式由柯西-施瓦茨不等式得出。因此，根据定义 ，

$$
\sigma_{\pi(k)}^2\geq(\Sigma-\Sigma_{\mathscr{E}})(\widehat{\mathbf{u}}_k^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)},\widehat{\mathbf{u}}_k^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(p)}) \tag{34}
$$

另一方面，

$$
\begin{aligned}
& (\Sigma-\Sigma_{\mathscr{E}})(\widehat{\mathbf{u}}_k^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)},\widehat{\mathbf{u}}_k^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(p)}) \\
\geq\ & \tilde{\Sigma}(\widehat{\mathbf{u}}_k^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)},\widehat{\mathbf{u}}_k^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(p)}) \\
& -\sup_{\|\mathbf{w}^{(q)}\|\leq 1,1\leq q\leq p}\left|(\widehat{\Sigma}-\Sigma)(\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w}^{(1)},\ldots,\widehat{\mathcal{P}}_\perp^{(p)}\mathbf{w}^{(p)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w}^{(1)},\ldots,\widehat{\mathcal{P}}_\perp^{(p)}\mathbf{w}^{(p)})\right| \\
\geq\ & \tilde{\Sigma}(\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_{l_*}^{(1)},\ldots,\widehat{\mathcal{P}}_\perp^{(p)}\mathbf{u}_{l_*}^{(p)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_{l_*}^{(1)},\ldots,\widehat{\mathcal{P}}_\perp^{(p)}\mathbf{u}_{l_*}^{(p)}) \\
& -\sup_{\|\mathbf{w}^{(q)}\|\leq 1,1\leq q\leq p}\left|(\widehat{\Sigma}-\Sigma)(\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w}^{(1)},\ldots,\widehat{\mathcal{P}}_\perp^{(p)}\mathbf{w}^{(p)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w}^{(1)},\ldots,\widehat{\mathcal{P}}_\perp^{(p)}\mathbf{w}^{(p)})\right| \\
\geq\ & (\Sigma-\Sigma_{\mathscr{E}})(\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_{l_*}^{(1)},\ldots,\widehat{\mathcal{P}}_\perp^{(p)}\mathbf{u}_{l_*}^{(p)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{u}_{l_*}^{(1)},\ldots,\widehat{\mathcal{P}}_\perp^{(p)}\mathbf{u}_{l_*}^{(p)}) \\
& -2\sup_{\|\mathbf{w}^{(q)}\|\leq 1,1\leq q\leq p}\left|(\widehat{\Sigma}-\Sigma)(\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w}^{(1)},\ldots,\widehat{\mathcal{P}}_\perp^{(p)}\mathbf{w}^{(p)},\widehat{\mathcal{P}}_\perp^{(1)}\mathbf{w}^{(1)},\ldots,\widehat{\mathcal{P}}_\perp^{(p)}\mathbf{w}^{(p)})\right|
\end{aligned}
$$

其中 l_* 是　　满足 $\sigma_{l_*}^2=\tau^2$的索引。按照之前相同的推导，我们有　　e

$$
(\Sigma-\Sigma_{\mathscr{E}})(\widehat{\mathbf{u}}_k^{(1)},\widehat{\mathbf{u}}_k^{(2)},\ldots,\widehat{\mathbf{u}}_k^{(p)},\widehat{\mathbf{u}}_k^{(1)},\ldots,\widehat{\mathbf{u}}_k^{(p)})\geq\tau^2\left(1-C\sqrt{\frac{r}{n}}-C\delta_k\right).
$$

结合（34），得到

$$
\sigma_{\pi(k)}^2\geq\tau^2\left(1-C\sqrt{\frac{r}{n}}-C\delta_k\right).
$$

结合 $\tilde{\Sigma}(\widehat{\mathrm{u}}_k^{(1)},\widehat{\mathrm{u}}_k^{(2)},\ldots,\widehat{\mathrm{u}}_k^{(p)},\widehat{\mathrm{u}}_k^{(1)},\ldots,\widehat{\mathrm{u}}_k^{(p)})$ 的下界和 $\tilde{\Sigma}(\widehat{\mathrm{P}}_\perp^{(1)}(\widehat{\mathrm{u}}_k^{(1)}-\mathrm{h}\widehat{\mathrm{u}}_{\pi_k}^{(1)},\mathrm{u}_{\pi_k}^{(1)})\mathrm{iu}_{\pi_k}^{(1)}))\ \widehat{\mathrm{u}}_k^{(2)},\ldots,\widehat{\mathrm{u}}_k^{(p)},\widehat{\mathrm{u}}_k^{(1)},\ldots,\widehat{\mathrm{u}}_k^{(p)})$ 的上界，我们有

$$
\begin{aligned}
\eta_k\ &\leq\ \frac{C\tau^2\delta_k^2+C(\sigma_0^2+\sigma_0\tau)\sqrt{\frac{d}{n}}}{\tau^2\left(1-C\delta_k^4-C\sqrt{\frac{r}{n}}\right)-C(\sigma_0\tau+\sigma_0^2)\sqrt{\frac{d}{n}}-C\sigma_0^2\delta_k^2} \\
&\leq\ C\delta_k^2+C\left(\frac{\sigma_0}{\tau}+\frac{\sigma_0^2}{\tau^2}\right)\sqrt{\frac{d}{n}} \\
&\leq\ C\left(\frac{\sigma_0}{\sigma_{\pi(k)}}+\frac{\sigma_0^2}{\sigma_{\pi(k)}^2}\right)\sqrt{\frac{d}{n}}=\delta_k.
\end{aligned}
$$

类似地，将 $\tilde{\Sigma}(\widehat{\mathrm{u}}_k^{(1)},\widehat{\mathrm{u}}_k^{(2)},\ldots,\widehat{\mathrm{u}}_k^{(p)},\widehat{\mathrm{u}}_k^{(1)},\ldots,\widehat{\mathrm{u}}_k^{(p)})$ 的下界与 $\max_{l\notin\pi([k])}\tilde{\Sigma}(\widehat{\mathrm{P}}_\perp^{(1)}\mathrm{u}_m^{(1)},\widehat{\mathrm{u}}_k^{(2)},\ldots,\widehat{\mathrm{u}}_k^{(p)},\widehat{\mathrm{u}}_k^{(1)},\ldots,\widehat{\mathrm{u}}_k^{(p)})$ 的上界结合起来，可得

$$
\begin{aligned}
\tilde{\eta}_k\ &\leq\ \frac{C\left(\tau^2\delta_k^2+(\tau^2\delta_k^{p-1}+\sigma_0\tau+\sigma_0^2)\sqrt{\frac{\log r}{n}}\right)}{\tau^2\left(1-C\delta_k^4-C\sqrt{\frac{r}{n}}\right)-C(\sigma_0\tau+\sigma_0^2)\sqrt{\frac{d}{n}}-C\sigma_0^2\delta_k^2} \\
&\leq\ C\delta_k^2+C\left(\delta_k+\frac{\sigma_0}{\tau}+\frac{\sigma_0^2}{\tau^2}\right)\sqrt{\frac{\log r}{n}} \\
&\leq\ C\delta_k^2+C\left(\frac{\sigma_0}{\sigma_{\pi(k)}}+\frac{\sigma_0^2}{\sigma_{\pi(k)}^2}\right)\sqrt{\frac{\log r}{n}}=\tilde{\delta}_k.
\end{aligned}
$$

C 定理的证明4.1和4.2

注意，定理 4.1可以看作定理 4.2 的特例，因此只需证明定理 4.2。与之前一样，我们只需考虑 $q = 1$ 的情形。记

$$\mathbf{w} = \langle \mathbf{w}, \mathbf{u}_{\pi(k)}^{(1)} \rangle \mathbf{u}_{\pi(k)}^{(1)} + \tilde{\mathbf{w}}.$$

Then

$$\langle \tilde{\mathbf{u}}_k^{(1)}, \mathbf{w} \rangle - \langle \mathbf{u}_{\pi(k)}^{(1)}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{u}_{\pi(k)}^{(1)} \rangle \left(\langle \tilde{\mathbf{u}}_k^{(1)}, \mathbf{u}_{\pi(k)}^{(1)} \rangle - 1 \right) + \langle \tilde{\mathbf{u}}_k^{(1)}, \tilde{\mathbf{w}} \rangle.$$

在假设 $d = o(n^{1/2})$下，根据引理1 和定理3.2，不难看出

$$\sin \angle (\tilde{\mathbf{u}}_k^{(1)}, \mathbf{u}_{\pi(k)}^{(1)}) = O_p \left(\sqrt{\frac{d}{n}} \right), \tag{35}$$

so

$$1 - \langle \tilde{\mathbf{u}}_k^{(1)}, \mathbf{u}_{\pi(k)}^{(1)} \rangle \leq 1 - \langle \tilde{\mathbf{u}}_k^{(1)}, \mathbf{u}_{\pi(k)}^{(1)} \rangle^2 = \sin^2 \angle (\tilde{\mathbf{u}}_k^{(1)}, \mathbf{u}_{\pi(k)}^{(1)}) = O_p \left(\frac{d}{n} \right) = o_p(n^{-1/2}).$$

因此，只需证明

$$\sqrt{n} \left(\langle \tilde{\mathbf{u}}_k^{(1)}, \tilde{\mathbf{w}} \rangle - \langle \mathbf{u}_{\pi(k)}^{(1)}, \tilde{\mathbf{w}} \rangle \right) \to_d N \left(0, \|\tilde{\mathbf{w}}\|^2 \left(\frac{\sigma_0^2}{\sigma_{\pi(k)}^2} + \frac{\sigma_0^4}{\sigma_{\pi(k)}^4} \right) \right).$$

回顾一下， $\tilde{}\mathbf{u}_k^{(1)}$ 是.....的主特征向量

$$\hat{\Sigma}(\cdot, \hat{\mathbf{u}}_k^{(2)}, \ldots, \hat{\mathbf{u}}_k^{(p)}, \cdot, \hat{\mathbf{u}}_k^{(2)}, \ldots, \hat{\mathbf{u}}_k^{(p)}),$$

这与.....的主特征向量相同

$$\tilde{\Sigma}(\cdot, \hat{\mathbf{u}}_k^{(2)}, \ldots, \hat{\mathbf{u}}_k^{(p)}, \cdot, \hat{\mathbf{u}}_k^{(2)}, \ldots, \hat{\mathbf{u}}_k^{(p)}),$$

so

$$\tilde{\mathbf{u}}_k^{(1)} = \frac{\tilde{\Sigma}(\cdot, \hat{\mathbf{u}}_k^{(2)}, \ldots, \hat{\mathbf{u}}_k^{(p)}, \tilde{\mathbf{u}}_k^{(1)}, \hat{\mathbf{u}}_k^{(2)}, \ldots, \hat{\mathbf{u}}_k^{(p)})}{\tilde{\Sigma}(\tilde{\mathbf{u}}_k^{(1)}, \hat{\mathbf{u}}_k^{(2)}, \ldots, \hat{\mathbf{u}}_k^{(p)}, \tilde{\mathbf{u}}_k^{(1)}, \hat{\mathbf{u}}_k^{(2)}, \ldots, \hat{\mathbf{u}}_k^{(p)})},$$

这意味着

$$\langle \tilde{\mathbf{u}}_k^{(1)}, \tilde{\mathbf{w}} \rangle = \frac{\tilde{\Sigma}(\tilde{\mathbf{w}}, \hat{\mathbf{u}}_k^{(2)}, \ldots, \hat{\mathbf{u}}_k^{(p)}, \tilde{\mathbf{u}}_k^{(1)}, \hat{\mathbf{u}}_k^{(2)}, \ldots, \hat{\mathbf{u}}_k^{(p)})}{\tilde{\Sigma}(\tilde{\mathbf{u}}_k^{(1)}, \hat{\mathbf{u}}_k^{(2)}, \ldots, \hat{\mathbf{u}}_k^{(p)}, \tilde{\mathbf{u}}_k^{(1)}, \hat{\mathbf{u}}_k^{(2)}, \ldots, \hat{\mathbf{u}}_k^{(p)})}.$$

我们先看分子。按照与子节（26）中相同的论证B.3，可得

$$\begin{aligned}&\tilde{\Sigma}(\tilde{\mathbf{w}}, \hat{\mathbf{u}}_k^{(2)}, \ldots, \hat{\mathbf{u}}_k^{(p)}, \tilde{\mathbf{u}}_k^{(1)}, \hat{\mathbf{u}}_k^{(2)}, \ldots, \hat{\mathbf{u}}_k^{(p)}) - \tilde{\Sigma}(\tilde{\mathbf{w}}, \mathbf{u}_{\pi(k)}^{(2)}, \ldots, \mathbf{u}_{\pi(k)}^{(p)}, \mathbf{u}_{\pi(k)}^{(1)}, \ldots, \mathbf{u}_{\pi(k)}^{(p)}) \\ \leq\, & C\left(\tau^2\delta_k^2 + \tau^2\delta^{p+1}\sqrt{\frac{r}{n}} + \tau^2\delta^{p-1}\sqrt{\frac{r}{n}} + \tau^2\delta_k^{2(p-1)} + \sigma_0\tau\delta_k\sqrt{\frac{d}{n}}\right) = o(n^{-1/2})\end{aligned}$$

以概率趋近于 1。因此

$$\begin{aligned}&\tilde{\Sigma}(\tilde{\mathbf{w}}, \hat{\mathbf{u}}_k^{(2)}, \ldots, \hat{\mathbf{u}}_k^{(p)}, \tilde{\mathbf{u}}_k^{(1)}, \hat{\mathbf{u}}_k^{(2)}, \ldots, \hat{\mathbf{u}}_k^{(p)}) \\ =&\tilde{\Sigma}(\tilde{\mathbf{w}}, \mathbf{u}_{\pi(k)}^{(2)}, \ldots, \mathbf{u}_{\pi(k)}^{(p)}, \mathbf{u}_{\pi(k)}^{(1)}, \ldots, \mathbf{u}_{\pi(k)}^{(p)}) + o_p(n^{-1/2}) \\ =&\sigma_{\pi(k)}\frac{1}{n}\sum_{i=1}^n \theta_{\pi(k)i}\mathscr{E}_i(\tilde{\mathbf{w}}, \mathbf{u}_{\pi(k)}^{(2)}, \ldots, \mathbf{u}_{\pi(k)}^{(p)}) \\ &+ \frac{1}{n}\sum_{i=1}^n \mathscr{E}_i(\mathbf{u}_{\pi(k)}^{(1)}, \ldots, \mathbf{u}_{\pi(k)}^{(p)})\mathscr{E}_i(\tilde{\mathbf{w}}, \mathbf{u}_{\pi(k)}^{(2)}, \ldots, \mathbf{u}_{\pi(k)}^{(p)}) - \mathbf{u}_{\pi(k)}^{(1)} + o_p(n^{-1/2}) \\ \to_d & N(0, (\sigma_0^4 + \sigma_0^2\sigma_{\pi(k)}^2)\|\tilde{\mathbf{w}}\|^2).\end{aligned}$$

另一方面，与上一节类似，

$$\tilde{\Sigma}(\tilde{\mathbf{u}}_k^{(1)}, \hat{\mathbf{u}}_k^{(2)}, \ldots, \hat{\mathbf{u}}_k^{(p)}, \tilde{\mathbf{u}}_k^{(1)}, \hat{\mathbf{u}}_k^{(2)}, \ldots, \hat{\mathbf{u}}_k^{(p)}) = \sigma_{\pi(k)}^2(1 + o_p(1)).$$

定理4.2 可由斯卢茨基定理得出。

定理4.1 中的断言

$$\begin{aligned}&\sqrt{n}\left[\mathrm{vec}(\tilde{\mathbf{U}}^{(q)}) - \mathrm{vec}(\mathbf{U}_\pi^{(q)})\right] \\ \xrightarrow{d}\, & N\left(0, \mathrm{diag}\left(\left(\frac{\sigma_0^2}{\sigma_{\pi(1)}^2} + \frac{\sigma_0^4}{\sigma_{\pi(1)}^4}\right)\mathcal{P}_{\mathbf{u}_{\pi(1)}^{(q)}}^{\perp}, \ldots, \left(\frac{\sigma_0^2}{\sigma_{\pi(r)}^2} + \frac{\sigma_0^4}{\sigma_{\pi(r)}^4}\right)\mathcal{P}_{\mathbf{u}_{\pi(r)}^{(q)}}^{\perp}\right)\right)\end{aligned}$$

可由定理 4.2 以及以下事实得出：

$$\sigma_{\pi(k)}\frac{1}{n}\sum_{i=1}^n \theta_{\pi(k)i}\mathscr{E}_i(\cdot, \mathbf{u}_{\pi(k)}^{(2)}, \ldots, \mathbf{u}_{\pi(k)}^{(p)}) + \frac{1}{n}\sum_{i=1}^n \mathscr{E}_i(\mathbf{u}_{\pi(k)}^{(1)}, \ldots, \mathbf{u}_{\pi(k)}^{(p)})\mathscr{E}_i(\cdot, \mathbf{u}_{\pi(k)}^{(2)}, \ldots, \mathbf{u}_{\pi(k)}^{(p)}) - \mathbf{u}_{\pi(k)}^{(1)}$$

对于任意 $k_1 \neq k_2$，它们都是独立的。

D 定理的证明4.3和4.4

定理 4.3 是定理 4.4 的一个特例，因此只需证明后者。我们首先需要引入一些记号。记

$$\bar{\mathbb{P}}(\cdot) := \mathbb{P}(\cdot|\mathscr{X}_{n/2+1}, \mathscr{X}_{n/2+2}, \ldots, \mathscr{X}_n), \tag{36}$$

and

$$\bar{\mathbb{E}}(\cdot) := \mathbb{E}(\cdot|\mathscr{X}_{n/2+1}, \mathscr{X}_{n/2+2}, \ldots, \mathscr{X}_n), \tag{37}$$

分别为给定$\{\mathscr{X}_{n/2+1}, \mathscr{X}_{n/2+2}, \ldots, \mathscr{X}_n\}$ 时的条件概率和期望。

令

$$\widehat{\mathcal{P}}_j^{(q)[1]} := \check{\mathbf{u}}_j^{(q)[1]} \otimes \check{\mathbf{u}}_j^{(q)[1]}, \qquad \mathcal{P}_j^{(q)} = \mathbf{u}_j^{(q)} \otimes \mathbf{u}_j^{(q)} = \mathcal{P}_{\mathbf{u}_j^{(q)}},$$

$$\mathcal{C}_j^{(q)} := \frac{1}{\sigma_j^2}\left(I_{d_q} - \mathbf{u}_j^{(q)} \otimes \mathbf{u}_j^{(q)}\right),$$

$$z_{jk} := \mathscr{X}_k \times_1 \widehat{\mathbf{u}}_j^{(1)[2]} \cdots \times_{q-1} \widehat{\mathbf{u}}_j^{(q-1)[2]} \times_{q+1} \widehat{\mathbf{u}}_j^{(q+1)[2]} \cdots \times_p \widehat{\mathbf{u}}_j^{(p)[2]}, \qquad k = 1, \ldots, n/2,$$

and

$$M_j^{(q)[1]} := \sum_{l=1}^{r} \sigma_l^2 \left(\prod_{q' \neq q} \langle \mathbf{u}_l^{(q')}, \widehat{\mathbf{u}}_j^{(q')[2]} \rangle\right)^2 \mathbf{u}_l^{(1)} \otimes \mathbf{u}_l^{(1)} + I_{d_q} = \sum_{l=1}^{r} \tilde{\sigma}_l^2 \mathbf{u}_l^{(1)} \otimes \mathbf{u}_l^{(1)} + I_{d_q},$$

其中

$$\tilde{\sigma}_l^2 = \sigma_l^2 \left(\prod_{q' \neq q} \langle \mathbf{u}_l^{(q')}, \widehat{\mathbf{u}}_j^{(q')[2]} \rangle\right)^2.$$

此外，令

$$\mathcal{C}_j^{(q)[1]} = \sum_{l \neq j} \frac{1}{\tilde{\sigma}_j^2 - \tilde{\sigma}_l^2} \mathbf{u}_l^{(q)} \otimes \mathbf{u}_l^{(q)} + \frac{1}{\tilde{\sigma}_j^2}\left(I_{d_q} - \sum_{l=1}^{r} \mathbf{u}_l^{(q)} \otimes \mathbf{u}_l^{(q)}\right),$$

$$\mathcal{L}_j^{(q)[1]} = \frac{2}{n} \sum_{k=1}^{n/2} \left(\mathcal{C}^{(q)[1]} z_{jk} \otimes \mathcal{P}_j^{(q)} z_{jk} + \mathcal{P}_j^{(q)} z_{jk} \otimes \mathcal{C}^{(q)[1]} z_{jk}\right),$$

$$\mathcal{S}_j^{(q)[1]} = \widehat{\mathcal{P}}_j^{(q)[1]} - \mathcal{P}_j^{(q)} - \mathcal{L}_j^{(q)[1]}.$$

我们使用花体大写字母$\mathcal{P}_j^{(q)}$，$\widehat{\mathcal{P}}_j^{(q)[1]}$、$\mathcal{C}_j^{(q)}$、$\mathcal{C}_j^{(q)[1]}$、$\mathcal{L}_j^{(q)[1]}$ 和$\mathcal{S}_j^{(q)[1]}$，以提醒读者它们是具有特定定义的矩阵。

定义

$$b_j^{(q)[1]} := \left\langle \bar{\mathbb{E}}(\mathcal{S}_j^{(q)[1]}) \mathbf{u}_j^{(q)}, \mathbf{u}_j^{(q)} \right\rangle. \tag{38}$$

并类似地定义 $b_j^{(q)[2]}$。最后，我们定义

$$b_j^{(q)} = \frac{\left\| \check{\mathbf{u}}_j^{(q),[1]} + \check{\mathbf{u}}_j^{(q),[2]} \right\|}{\sqrt{1 + b_j^{(q)[1]}} + \sqrt{1 + b_j^{(q)[2]}}} - 1.. \tag{39}$$

证明过程相当复杂，我们将其分成几个步骤。

步骤 1. 我们将把 $\check{\mathbf{u}}_j^{(q)\,[1]}$ 的线性形式表示为 $\hat{\mathcal{P}}_j^{(q)[1]}$ 的双线性形式，并证明

$$\sqrt{n}\left((1+b_j^{(q)})\langle\check{\mathbf{u}}_j^{(q)},\mathbf{v}\rangle-\langle\mathbf{u}_{\pi(j)}^{(q)},\mathbf{v}\rangle\right)\to_d N\left(0,\left(\frac{\sigma_0^2}{\sigma_{\pi(j)}^2}+\frac{\sigma_0^4}{\sigma_{\pi(j)}^4}\right)\|\mathcal{P}_{\mathbf{u}_{\pi(j)}^{(q)}}^{\perp}\mathbf{v}\|^2\right). \tag{40}$$

步骤 2。 我们证明

$$\langle\check{\mathbf{u}}_j^{(q)},\mathbf{u}_{\pi(j)}^{(q)}\rangle=\frac{1}{\sqrt{1+\frac{d_q}{n}\left(\frac{\sigma_0^2}{\sigma_{\pi(j)}^2}+\frac{\sigma_0^4}{\sigma_{\pi(j)}^4}\right)}}+O_p\left(\frac{d^{3/2}}{n^{3/2}}\right). \tag{41}$$

注意，令 $\mathbf{v}=\mathbf{u}_{\pi(j)}^{(q)}$，则（40）蕴含

$$(1+b_j^{(q)})\langle\check{\mathbf{u}}_j^{(q)},\mathbf{u}_{\pi(j)}^{(q)}\rangle-1=o_p(n^{-1/2}).$$

将其与（41）结合，我们立即得到

$$b_j^{(q)}=\sqrt{1+\frac{d_q}{n}\left(\frac{\sigma_0^2}{\sigma_{\pi(j)}^2}+\frac{\sigma_0^4}{\sigma_{\pi(j)}^4}\right)}-1+O_p\left(\frac{d^{3/2}}{n^{3/2}}\right)+o_p\left(\frac{1}{\sqrt{n}}\right),$$

步骤 3。 最后，我们证明

$$\sqrt{n}\left(\widehat{b}_j^{(q)}-b_j^{(q)}\right)\xrightarrow{p}0. \tag{42}$$

为简单起见，在证明的其余部分中，不失一般性地假设，将 π 与 $\hat{\mathbf{u}}_j^{(q)}$ 匹配到 $\mathbf{u}_{\pi(j)}^{(1)}$ 的置换是恒等置换。我们还将反复使用以下事实，通常不作明确说明。

类似于定理3.2的证明，记

$$\widehat{\Sigma}_{\mathscr{E}}^{[1]}=\frac{2}{n}\sum_{i=1}^{n/2}\mathscr{E}_i\otimes\mathscr{E}_i,\qquad\widehat{\Sigma}_{\theta,\mathscr{E}}^{[1]}=\frac{2}{n}\sum_{i=1}^{n/2}\theta_i\otimes\mathscr{E}_i.$$

根据引理1 和2，其概率趋于 1，

$$\sigma_0^{-2}\|\widehat{\Sigma}_{\mathscr{E}}^{[1]}-\mathscr{I}\|\le C\sqrt{\frac{d}{n}},\ \sigma_0^{-1}\|\widehat{\Sigma}_{\theta,\mathscr{E}}^{[1]}\|\le C\sqrt{\frac{d}{n}}. \tag{43}$$

Let

$$\Delta^{(q),[2]}=\hat{\mathbf{u}}_j^{(q),[2]}-\mathbf{u}_j^{(q)}.$$

根据定理3.2，我们有

$$\max_{q\in[p]}\|\Delta^{(q),[2]}\| = O_p\left(\sqrt{\frac{d}{n}}\right)$$

此外，在假设 $d = o(n)$ 下，根据引理 1 和定理 3.2，不难看出

$$\sin\angle(\check{\mathbf{u}}_j^{(1)[1]}, \mathbf{u}_j^{(1)}) = O_p\left(\sqrt{\frac{d}{n}}\right).$$

记

$$\delta^{(1)[2]} := \max\left\{\max_{q\in[p]}\|\Delta^{(q),[2]}\|, \|\check{\mathbf{u}}_j^{(1)[1]} - \mathbf{u}_j^{(1)}\|\right\},$$

结合上述两个 界，我们可知，在定理4.4的假设下 ，

$$\delta^{(1)[2]} = O_p\left(\sqrt{\frac{d}{n}}\right). \tag{44}$$

D.1 步骤 1.

不失一般性，在这一步中我们假设 $\sigma_0 = 1$。我们只需证明$q = 1$的情形，因此在这一步中，为简便起见，我们也将 $\mathcal{P}_j = \mathcal{P}_j^{(1)} = \mathcal{P}^{\mathbf{u}_j^{(1)}}$ 和 $\mathcal{C}_j = \mathcal{C}_j^{(1)} = \frac{1}{\sigma_j^2}(I - \mathcal{P}_j)$ 记作。 —

定义

$$\eta_{jk}(\mathbf{v}) := \langle y_{jk}, \mathcal{P}_j\mathbf{u}_j^{(1)}\rangle\langle y_{jk}, \mathcal{C}_j\mathbf{v}\rangle,$$

其中

$$y_{jk} := \mathscr{X}_k \times_2 \mathbf{u}_j^{(2)} \cdots \times_p \mathbf{u}_j^{(p)} = \sigma_j\theta_{jk}\mathbf{u}_j^{(1)} + \mathscr{E}_k \times_2 \mathbf{u}_j^{(2)} \cdots \times_p \mathbf{u}_j^{(p)}.$$

回顾一下，$\widehat{\mathcal{P}}_j^{(1)[1]} := \check{\mathbf{u}}_j^{(1)[1]} \otimes \check{\mathbf{u}}_j^{(1)[1]}$，以及 $\widehat{\mathcal{P}}_j^{(1)[2]} := \check{\mathbf{u}}_j^{(1)[2]} \otimes \check{\mathbf{u}}_j^{(1)[2]}$。定义

$$\rho_j(\mathbf{v})^{[1]} := \left\langle((\widehat{\mathcal{P}}_j^{(1)[1]} - (1 + b_j^{(1)[1]})\mathcal{P}_j)\mathbf{u}_j^{(1)}, \mathbf{v}\right\rangle, \mathbf{v} \in \mathbb{H}. \tag{45}$$

$$\rho_j(\mathbf{v})^{[2]} := \left\langle((\widehat{\mathcal{P}}_j^{[2]} - (1 + b_j^{(1)[2]})\mathcal{P}_j)\mathbf{u}_j^{(1)}, \mathbf{v}\right\rangle, \mathbf{v} \in \mathbb{H}. \tag{46}$$

Koltchinskii 和 Lounici （2014）中的方程（6.6）给出了 $\check{\mathbf{u}}_j^{(1)[1]}$,的线性形式关于 $\rho_j(\mathbf{v})^{[1]}$ 和 $\rho_j(\mathbf{v})^{[2]}$的表示:

$$\begin{aligned}
&\sqrt{n}\left\langle \frac{1}{2}\left[\check{\mathbf{u}}_j^{(1),[1]}+\check{\mathbf{u}}_j^{(1),[2]}-\left(\sqrt{1+b_j^{(1)[1]}}+\sqrt{1+b_j^{(1)[2]}}\right)\mathbf{u}_j^{(1)}\right],\mathbf{v}\right\rangle \\
=&\sum_{h=1}^{2}\sqrt{n}\Bigg[\frac{\rho_j(\mathbf{v})^{[h]}}{\sqrt{1+b_j^{(1)[h]}+\rho_j(\mathbf{v})^{[h]}}} \\
&+\frac{\sqrt{1+b_j^{(1)[h]}}}{\sqrt{1+b_j^{(1)[h]}+\rho_j(\mathbf{v})^{[h]}}\left(\sqrt{1+b_j^{(1)[h]}+\rho_j(\mathbf{v})^{[h]}}+\sqrt{1+b_j^{(1)[h]}}\right)}\rho_j(u_j^{(1)})^{[h]}\langle\mathbf{u}_j^{(1)},\mathbf{v}\rangle\Bigg]. \qquad (47)
\end{aligned}$$

我们将用到以下引理:

引理 3. 在定理 4.4的相同假设下，对于任意 $\mathbf{u}$，$\mathbf{v}\in\mathbb{R}^{d_q}$,

$$\begin{aligned}
&\sqrt{\frac{n}{2}}\left\langle\left[\widehat{\mathcal{P}}_j^{(q)[1]}-(1+b_j^{(q)[1]})\mathcal{P}_j^{(q)}\right]\mathbf{u},\mathbf{v}\right\rangle \\
&-\sqrt{\frac{2}{n}}\sum_{k=1}^{n/2}\left[\langle y_{jk}^{(q)},\mathcal{P}_j^{(q)}\mathbf{v}\rangle\langle y_{jk}^{(q)},\mathcal{C}_j^{(q)}\mathbf{u}\rangle+\langle y_{jk}^{(q)},\mathcal{P}_j^{(q)}\mathbf{u}\rangle\langle y_{jk}^{(q)},\mathcal{C}_j^{(q)}\mathbf{v}\rangle\right]=o_p\left(\|\mathbf{u}\|\|\mathbf{v}\|\right).
\end{aligned}$$

此外，存在普适常数 C 使得 $\mathbb{P}(|b_j^{(q)[1]}|\le Cd/n)\to 1$ 当 $n\to\infty$时。

由引理 3可知，注意到 $\mathcal{C}j\mathbf{u}_j^{(1)}=0$，我们有

$$\sqrt{\frac{n}{2}}\rho_j(\mathbf{v})^{[1]}-\sqrt{\frac{2}{n}}\sum_{k=1}^{\frac{n}{2}}\eta_{jk}(\mathbf{v})\xrightarrow{p}0, \qquad (48)$$

$$\sqrt{\frac{n}{2}}\rho_j(\mathbf{v})^{[2]}-\sqrt{\frac{2}{n}}\sum_{k=\frac{n}{2}+1}^{n}\eta_{jk}(\mathbf{v})\xrightarrow{p}0. \qquad (49)$$

如果我们能够证明对于任意 $\mathbf{v}\in\mathbb{R}^{d_1}$,

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\eta_{jk}(\mathbf{v})\xrightarrow{d}N\left(0,\frac{1+\sigma_j^2}{\sigma_j^4}\|\mathcal{P}_j^{\perp}\mathbf{v}\|^2\right), \qquad (50)$$

然后结合以下事实： $b_j^{(1)[h]}\xrightarrow{p}0$ 以及（47）、（48）和（49，可得:

$$\begin{aligned}
&\sqrt{n}\left\langle\frac{1}{2}\left[\check{\mathbf{u}}_j^{(1),[1]}+\check{\mathbf{u}}_j^{(1),[2]}-\left(\sqrt{1+b_j^{(1)[1]}}+\sqrt{1+b_j^{(1)[2]}}\right)\mathbf{u}_j^{(1)}\right],\mathbf{v}\right\rangle \\
=&\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\eta_{jk}(\mathbf{v})+o_p(1) \\
\xrightarrow{d}&N\left(0,\frac{1+\sigma_j^2}{\sigma_j^4}\|\mathcal{P}_j^{\perp}\mathbf{v}\|^2\right).
\end{aligned}$$

回忆一下， $b_j^{(1)[h]} \xrightarrow{p} 0$。根据斯卢茨基定理，

$$\begin{aligned}
&\sqrt{n}\left((1+b_j^{(1)})\langle \check{\mathbf{u}}_j^{(1)}, \mathbf{v}\rangle - \langle \mathbf{u}_j^{(1)}, \mathbf{v}\rangle\right) \\
=&\frac{1}{\sqrt{1+b_j^{(1)[1]}}+\sqrt{1+b_j^{(1)[2]}}} \cdot \sqrt{n}\left\langle \frac{1}{2}\left[\check{\mathbf{u}}_j^{(1),[1]} + \check{\mathbf{u}}_j^{(1),[2]} - \left(\sqrt{1+b_j^{(1)[1]}}+\sqrt{1+b_j^{(1)[2]}}\right)\mathbf{u}_j^{(1)}\right], \mathbf{v}\right\rangle \\
\xrightarrow{d} &N\left(0, \frac{1+\sigma_j^2}{\sigma_j^4}\|\mathcal{P}_j^{\perp}\mathbf{v}\|^2\right).
\end{aligned}$$

命题（40）随即得证。

现在我们来证明（50）。注意到 $\mathcal{P}_j\mathbf{u}_j^{(1)} = \mathbf{u}_j^{(1)}$，且 $\mathcal{C}_j = \frac{1}{\sigma_j^2}\mathcal{P}_j^{\perp}$，我们有

$$\eta_{jk}(\mathbf{v}) = \frac{1}{\sigma_j^2}\left(\sigma_j\theta_{jk} + \mathscr{E}_k \times_1 \mathbf{u}_j^{(1)} \times_2 \mathbf{u}_j^{(2)} \cdots \times_p \mathbf{u}_j^{(p)}\right)\left(\mathscr{E}_k \times_1 (\mathcal{P}_j^{\perp}\mathbf{v}) \times_2 \mathbf{u}_j^{(2)} \cdots \times_p \mathbf{u}_j^{(p)}\right).$$

两者 $\sigma_j\theta_{jk} + \mathscr{E}_k \times_1 \mathbf{u}_j^{(1)} \times_2 \mathbf{u}_j^{(2)} \cdots \times_p \mathbf{u}_j^{(p)}$ 和 $\mathscr{E}_k \times_1 (\mathbb{I}_{d_1}\mathcal{P}_j^{\perp}\mathbf{v}) \times_2 \mathbf{u}_j^{(2)} \cdots \times_p \mathbf{u}_j^{(p)}$ 都是中心高斯随机变量。此外，它们相互独立，因为 θ_{jk} 与$\mathscr{E}_k$独立，且 $\mathbf{u}_j^{(1)} \perp \mathcal{P}_j^{\perp}\mathbf{v}$。因此

$$\mathbb{E}\eta_{jk}(\mathbf{v}) = 0.$$

更一般地，高斯变量 $\sigma_j\theta_{jk} + \mathscr{E}_k \times_1 \mathbf{u}_j^{(1)} \times_2 \mathbf{u}_j^{(2)} \cdots \times_p \mathbf{u}_j^{(p)}$ 和高斯向量$\mathscr{E}_k \times_1 (\mathcal{P}_j^{\perp}) \times_2 \mathbf{u}_j^{(2)} \cdots \times_p \mathbf{u}_j^{(p)}$ 出于完全相同的原因而相互独立。因此，直接计算可得：

$$\begin{aligned}
&\mathrm{var}(\eta_{jk}(\mathbf{v})) \\
=&\frac{1}{\sigma_j^4}\mathbb{E}\left(\sigma_j\theta_{jk} + \mathscr{E}_k \times_1 \mathbf{u}_j^{(1)} \times_2 \mathbf{u}_j^{(2)} \cdots \times_p \mathbf{u}_j^{(p)}\right)^2 \\
&\mathbb{E}\left[\left(\mathscr{E}_k \times_1 (\mathcal{P}_j^{\perp}\mathbf{v}) \times_2 \mathbf{u}_j^{(2)} \cdots \times_p \mathbf{u}_j^{(p)}\right)^2\right] \\
=&\frac{1+\sigma_j^2}{\sigma_j^4}\|\mathcal{P}_j^{\perp}\mathbf{v}\|^2.
\end{aligned}$$

鉴于对于 $k = 1, 2, \ldots, n$， $\eta_{jk}(\mathbf{v})$ 是独立同分布的，要利用中心极限定理完成这部分证明，还需验证中心极限定理的 Lindeberg条件，这可归结为

$$\frac{\mathbb{E}\eta_{jk}(\mathbf{v})^2\mathbb{I}\left(|\eta_{jk}(\mathbf{v})| \geq \tau\sqrt{n}\mathbb{E}^{1/2}\eta_{jk}(\mathbf{v})^2\right)}{\mathbb{E}\eta_{jk}(\mathbf{v})^2} \to 0 \text{ as } n \to \infty$$

对于所有 $\tau > 0$。注意，

$$\frac{\mathbb{E}\eta_{jk}(\mathbf{v})^2\mathbb{I}\left(|\eta_{jk}(\mathbf{v})| \geq \tau\sqrt{n}\mathbb{E}^{1/2}\eta_{jk}(\mathbf{v})^2\right)}{\mathbb{E}\eta_{jk}(\mathbf{v})^2} \leq \frac{\mathbb{E}\eta_{jk}(\mathbf{v})^4}{\tau^2 n\left(\mathbb{E}\eta_{jk}(\mathbf{v})^2\right)^2}.$$

由于

$$\begin{aligned}&\mathbb{E}\eta_{jk}(\mathbf{v})^4\\=&\mathbb{E}\left(\sigma_j\theta_{jk}+\mathscr{E}_k\times_1\mathbf{u}_j^{(1)}\times_2\mathbf{u}_j^{(2)}\cdots\times_p\mathbf{u}_j^{(p)}\right)^4\mathbb{E}\left(\mathscr{E}_k\times_1(\mathcal{P}_j^{\perp}\mathbf{v})\times_2\mathbf{u}_j^{(2)}\cdots\times_p\mathbf{u}_j^{(p)}\right)^4,\end{aligned}$$

and

$$\begin{aligned}&\left(\mathbb{E}\eta_{jk}(\mathbf{v})^2\right)^2\\=&\mathbb{E}^2\left(\sigma_j\theta_{jk}+\mathscr{E}_k\times_1\mathbf{u}_j^{(1)}\times_2\mathbf{u}_j^{(2)}\cdots\times_p\mathbf{u}_j^{(p)}\right)^2\mathbb{E}^2\left(\mathscr{E}_k\times_1(\mathcal{P}_j^{\perp}\mathbf{v})\times_2\mathbf{u}_j^{(2)}\cdots\times_p\mathbf{u}_j^{(p)}\right)^2,\end{aligned}$$

根据以下事实：对于中心化正态随机变量 ξ, $\mathbb{E}\xi^4=3\mathbb{E}^2\xi^2$，我们得到

$$\frac{\mathbb{E}\eta_{jk}(\mathbf{v})^4}{\tau^2n\left(\mathbb{E}\eta_{jk}(\mathbf{v})^2\right)^2}=\frac{1}{\tau^2n}\to0,$$

并且（50）得证。

D.2 第 2 步。

同样，只需考虑情形 $q=1$。现在我们将证明

$$\langle\check{\mathbf{u}}_j^{(1)},\mathbf{u}_j^{(1)}\rangle=\frac{1}{\sqrt{1+\frac{d_1}{n}\left(\frac{\sigma_0^2}{\sigma_j^2}+\frac{\sigma_0^4}{\sigma_j^4}\right)}}+O_p\left(\frac{d^{3/2}}{n^{3/2}}\right).\tag{51}$$

记 $\tilde{\Sigma}^{[1]}=\widehat{\Sigma}^{[1]}-\mathscr{I}$, $\tilde{\Sigma}^{[2]}=\widehat{\Sigma}^{[2]}-\mathscr{I}$，并令

$$\grave{\mathbf{u}}_j^{(1),[1]}=\tilde{\Sigma}^{[1]}(\cdot,\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]},\check{\mathbf{u}}_j^{(1),[1]},\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]}),$$

and

$$\grave{\mathbf{u}}_j^{(1),[2]}=\tilde{\Sigma}^{[2]}(\cdot,\widehat{\mathbf{u}}_j^{(2)[1]},\ldots,\widehat{\mathbf{u}}_j^{(p)[1]},\check{\mathbf{u}}_j^{(1),[2]},\widehat{\mathbf{u}}_j^{(2)[1]},\ldots,\widehat{\mathbf{u}}_j^{(p)[1]}).$$

由于 $\check{}\mathbf{u}_j^{(1)\,[1]}$,是.....的主特征向量

$$\widehat{\Sigma}^{[1]}(\cdot,\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]},\cdot,\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]}),$$

它也是.....的主特征向量

$$\tilde{\Sigma}^{[1]}(\cdot,\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]},\cdot,\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]}),$$

so

$$\check{\mathbf{u}}_j^{(1),[1]} = \frac{\grave{\mathbf{u}}_j^{(1),[1]}}{\|\grave{\mathbf{u}}_j^{(1),[1]}\|},$$

类似地

$$\check{\mathbf{u}}_j^{(1),[2]} = \frac{\grave{\mathbf{u}}_j^{(1),[2]}}{\|\grave{\mathbf{u}}_j^{(1),[2]}\|}.$$

写出

$$\mathbf{z}_j^{(1)[1]} := \grave{\mathbf{u}}_j^{(1),[1]} - \langle \grave{\mathbf{u}}_j^{(1),[1]}, \mathbf{u}_j^{(1)} \rangle \mathbf{u}_j^{(1)} - \mathbf{y}_j^{(1)[1]},$$

其中

$$\mathbf{y}_j^{(1)[1]} := \mathcal{P}_{\mathbf{u}_j^{(1)}}^{\perp} \left[\frac{2}{n} \sum_{k=1}^{n/2} \mathscr{E}_k(\cdot, \mathbf{u}_j^{(2)}, \ldots, \mathbf{u}_j^{(p)}) \sigma_j \theta_{jk} + \frac{2}{n} \sum_{k=1}^{n/2} \mathscr{E}_k(\cdot, \mathbf{u}_j^{(2)}, \ldots, \mathbf{u}_j^{(p)}) \mathscr{E}_k(\mathbf{u}_j^{(1)}, \ldots, \mathbf{u}_j^{(p)}) \right]. \tag{52}$$

$\mathbf{z}_j^{(1)[2]}$ 和 $\mathbf{y}_j^{(1)[2]}$ 以类似方式定义，使得

$$\grave{\mathbf{u}}_j^{(1),[1]} = \langle \grave{\mathbf{u}}_j^{(1),[1]}, \mathbf{u}_j^{(1)} \rangle \mathbf{u}_j^{(1)} + \mathbf{y}_j^{(1)[1]} + \mathbf{z}_j^{(1)[1]},$$

$$\grave{\mathbf{u}}_j^{(1),[2]} = \langle \grave{\mathbf{u}}_j^{(1),[2]}, \mathbf{u}_j^{(1)} \rangle \mathbf{u}_j^{(1)} + \mathbf{y}_j^{(1)[2]} + \mathbf{z}_j^{(1)[2]}.$$

我们将分别处理 $\grave{\mathbf{u}}_j^{(1)\,[1]}$,和 $\grave{\mathbf{u}}_j^{(1)\,[1]}$,这三项，以证明

$$\langle \grave{\mathbf{u}}_j^{(1),[1]}, \mathbf{u}_j^{(1)} \rangle = \sigma_j^2 + O_p\left(\frac{d}{n}\right), \ \langle \grave{\mathbf{u}}_j^{(1),[2]}, \mathbf{u}_j^{(1)} \rangle = \sigma_j^2 + O_p\left(\frac{d}{n}\right), \tag{53}$$

$$\left\| \mathbf{y}_j^{(1)[1]} + \mathbf{y}_j^{(1)[2]} \right\| = 2\sqrt{\frac{d_1}{n}} \sqrt{\sigma_0^2 \sigma_j^2 + \sigma_0^4} + O_p\left(\frac{\sqrt{d}}{n}\right) + O_p\left(\frac{1}{\sqrt{n}}\right), \tag{54}$$

$$\begin{aligned} \left\| \mathbf{y}_j^{(1)[1]} \right\| =& \sqrt{\frac{2d_1}{n}} \sqrt{\sigma_0^2 \sigma_j^2 + \sigma_0^4} + O_p\left(\frac{\sqrt{d}}{n}\right) + O_p\left(\frac{1}{\sqrt{n}}\right), \\ \left\| \mathbf{y}_j^{(1)[2]} \right\| =& \sqrt{\frac{2d_1}{n}} \sqrt{\sigma_0^2 \sigma_j^2 + \sigma_0^4} + O_p\left(\frac{\sqrt{d}}{n}\right) + O_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned} \tag{55}$$

and

$$\mathbf{z}_j^{(1)[1]} = O_p\left(\frac{d}{n}\right), \ \mathbf{z}_j^{(1)[2]} = O_p\left(\frac{d}{n}\right). \tag{56}$$

我们首先证明，方程（51）可由（53）-（56）中的界推出。我们从 $\|\grave{\mathbf{u}}_j^{(1)[1]},\|$ 开始。结合（53）-（56），我们有

$$\begin{aligned}
&\|\grave{\mathbf{u}}_j^{(1),[1]}\| \\
=&\sqrt{\left\langle \grave{\mathbf{u}}_j^{(1),[1]}, \mathbf{u}_j^{(1)} \right\rangle^2 + \left\| \mathbf{y}_j^{(1)[1]} + \mathbf{z}_j^{(1)[1]} \right\|^2} \\
=&\sqrt{\left[\sigma_j^2 + O_p\left(\frac{d}{n}\right)\right]^2 + \left[O_p\left(\sqrt{\frac{d}{n}}\right) + O_p\left(\frac{d}{n}\right)\right]^2} \\
=&\sigma_j^2 + O_p\left(\frac{d}{n}\right),
\end{aligned} \tag{57}$$

其中，第二个等式由（53）、（55）、（56）可得，而最后一个等式成立是因为$d = o(n)$。类似地，我们可以推导出

$$\|\grave{\mathbf{u}}_j^{(1),[2]}\| = \sigma_j^2 + O_p\left(d/n\right).$$

因此，

$$\begin{aligned}
&\left\| \frac{\mathbf{y}_j^{(1)[1]}}{\|\grave{\mathbf{u}}_j^{(1),[1]}\|} + \frac{\mathbf{y}_j^{(1)[2]}}{\|\grave{\mathbf{u}}_j^{(1),[2]}\|} \right\| \\
=&\left\| \frac{1}{\sigma_j^2}(\mathbf{y}_j^{(1)[1]} + \mathbf{y}_j^{(1)[2]}) + \left(\frac{1}{\|\grave{\mathbf{u}}_j^{(1),[1]}\|} - \frac{1}{\sigma_j^2}\right)\mathbf{y}_j^{(1)[1]} + \left(\frac{1}{\|\grave{\mathbf{u}}_j^{(1),[2]}\|} - \frac{1}{\sigma_j^2}\right)\mathbf{y}_j^{(1)[2]} \right\| \\
=&2\sqrt{\frac{d_1}{n}\left(\frac{\sigma_0^2}{\sigma_j^2} + \frac{\sigma_0^4}{\sigma_j^4}\right)} + O_p\left(\frac{\sqrt{d}}{n}\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{\sqrt{d}}{n}\right) \cdot O_p\left(\frac{1}{\sqrt{n}}\right) \\
=&2\sqrt{\frac{d_1}{n}\left(\frac{\sigma_0^2}{\sigma_j^2} + \frac{\sigma_0^4}{\sigma_j^4}\right)} + O_p\left(\frac{\sqrt{d}}{n}\right) + O_p\left(\frac{1}{\sqrt{n}}\right),
\end{aligned} \tag{58}$$

其中，第二个等式由（54）、（55）和（57）可得。

此外，（56）、（57）意味着

$$\left\| \frac{\mathbf{z}_j^{(1)[1]}}{\|\grave{\mathbf{u}}_j^{(1),[1]}\|} + \frac{\mathbf{z}_j^{(1)[2]}}{\|\grave{\mathbf{u}}_j^{(1),[2]}\|} \right\| = O_p\left(\frac{d}{n}\right), \tag{59}$$

且（53）和（57）意味着

$$\frac{\langle \grave{\mathbf{u}}_j^{(1),[1]}, \mathbf{u}_j^{(1)} \rangle}{\|\grave{\mathbf{u}}_j^{(1),[1]}\|} + \frac{\langle \grave{\mathbf{u}}_j^{(1),[2]}, \mathbf{u}_j^{(1)} \rangle}{\|\grave{\mathbf{u}}_j^{(1),[2]}\|} = 2 + O_p\left(\frac{d}{n}\right). \tag{60}$$

因此，

$$
\begin{aligned}
&\langle \check{\mathbf{u}}_j^{(1)}, \mathbf{u}_j^{(1)} \rangle \\
=&\frac{\frac{\langle \grave{\mathbf{u}}_j^{(1),[1]}, \mathbf{u}_j^{(1)} \rangle}{\|\grave{\mathbf{u}}_j^{(1),[1]}\|} + \frac{\langle \grave{\mathbf{u}}_j^{(1),[2]}, \mathbf{u}_j^{(1)} \rangle}{\|\grave{\mathbf{u}}_j^{(1),[2]}\|}}{\sqrt{\left(\frac{\langle \grave{\mathbf{u}}_j^{(1),[1]}, \mathbf{u}_j^{(1)} \rangle}{\|\grave{\mathbf{u}}_j^{(1),[1]}\|} + \frac{\langle \grave{\mathbf{u}}_j^{(1),[2]}, \mathbf{u}_j^{(1)} \rangle}{\|\grave{\mathbf{u}}_j^{(1),[2]}\|} \right)^2 + \left(\frac{\mathbf{y}_j^{(1)[1]}}{\|\grave{\mathbf{u}}_j^{(1),[1]}\|} + \frac{\mathbf{y}_j^{(1)[2]}}{\|\grave{\mathbf{u}}_j^{(1),[2]}\|} + \frac{\mathbf{z}_j^{(1)[1]}}{\|\grave{\mathbf{u}}_j^{(1),[1]}\|} + \frac{\mathbf{z}_j^{(1)[2]}}{\|\grave{\mathbf{u}}_j^{(1),[2]}\|} \right)^2}} \\
=&1 \Bigg/ \sqrt{1 + \left(\frac{\frac{\mathbf{y}_j^{(1)[1]}}{\|\grave{\mathbf{u}}_j^{(1),[1]}\|} + \frac{\mathbf{y}_j^{(1)[2]}}{\|\grave{\mathbf{u}}_j^{(1),[2]}\|} + \frac{\mathbf{z}_j^{(1)[1]}}{\|\grave{\mathbf{u}}_j^{(1),[1]}\|} + \frac{\mathbf{z}_j^{(1)[2]}}{\|\grave{\mathbf{u}}_j^{(1),[2]}\|}}{\frac{\langle \grave{\mathbf{u}}_j^{(1),[1]}, \mathbf{u}_j^{(1)} \rangle}{\|\grave{\mathbf{u}}_j^{(1),[1]}\|} + \frac{\langle \grave{\mathbf{u}}_j^{(1),[2]}, \mathbf{u}_j^{(1)} \rangle}{\|\grave{\mathbf{u}}_j^{(1),[2]}\|}} \right)^2} \\
=&1 \Bigg/ \sqrt{1 + \left(\frac{2\sqrt{\frac{d_1}{n}\left(\frac{\sigma_0^2}{\sigma_j^2} + \frac{\sigma_0^4}{\sigma_j^4} \right)} + O_p\left(\frac{\sqrt{d}}{n} \right) + O_p\left(\frac{1}{\sqrt{n}} \right) + O_p\left(\frac{d}{n} \right)}{2 + O_p\left(\frac{d}{n} \right)} \right)^2} \\
=&1 \Bigg/ \sqrt{1 + \left\{ \left[\sqrt{\frac{d_1}{n}\left(\frac{\sigma_0^2}{\sigma_j^2} + \frac{\sigma_0^4}{\sigma_j^4} \right)} + O_p\left(\frac{1}{\sqrt{n}} \right) + O_p\left(\frac{d}{n} \right) \right] \cdot \left[1 + O_p\left(\frac{d}{n} \right) \right] \right\}^2} \\
=&\frac{1}{\sqrt{1 + \left[\sqrt{\frac{d_1}{n}\left(\frac{\sigma_0^2}{\sigma_j^2} + \frac{\sigma_0^4}{\sigma_j^4} \right)} + O_p\left(\frac{d}{n} \right) \right]^2}} \\
=&\frac{1}{\sqrt{1 + \frac{d_1}{n}\left(\frac{\sigma_0^2}{\sigma_j^2} + \frac{\sigma_0^4}{\sigma_j^4} \right)}} + O_p\left(\frac{d^{3/2}}{n^{3/2}} \right),
\end{aligned}
$$

其中第三个方程由（58）-（60）得出，而最后一个方程则由以下简单事实得出

$$
\left| \frac{1}{\sqrt{1 + (x+y)^2}} - \frac{1}{\sqrt{1 + x^2}} \right| \leq \left| xy + \frac{1}{2} y^2 \right|.
$$

现在只需证明（53）-（56）。

D.2.1 方程（54）和（55）。

写出

$$
\mathbf{x}_k := \mathscr{E}_k(\cdot, \mathbf{u}_j^{(2)}, \ldots, \mathbf{u}_j^{(p)}) \in \mathbb{R}^{d_1}.
$$

如果我们选择 $\mathbb{R}^{d_1}$ 的一组正交基，其中 $\mathbf{e}_1 = \mathbf{u}_j^{(1)}$ ，那么

$$\left\|\frac{1}{2}(\mathbf{y}_j^{(1)[1]} + \mathbf{y}_j^{(1)[2]})\right\| = \left\|\mathcal{P}^{\perp}_{\mathbf{u}_j^{(1)}}\left[\frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k\Big(\sigma_j\theta_{jk} + \mathbf{x}_{1,k}\Big)\right]\right\| = \left\|\frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_{(-1),k}\Big(\sigma_j\theta_{jk} + \mathbf{x}_{1,k}\Big)\right\|,$$

其中 $\mathbf{x}_{1,k}$ 表示$\mathbf{x}_k$的第一个分量，而 $\mathbf{x}_{(-1),k}$ 表示所有其他分量（维数为 $d_1 - 1$）。

注意，$\mathbf{x}_{(-1),k}$和 $\Big(\sigma_j\theta_{jk} + \mathbf{x}_{1,k}\Big)$是两组相互独立的独立同分布随机变量，$\mathbf{x}_{(-1),k}$服从分布 $N(\mathbf{0}, \sigma_0^2 I_{d_1-1})$，而 $\Big(\sigma_j\theta_{jk} + \mathbf{x}_{1,k}\Big)$服从分布 $N(\mathbf{0}, \sigma_0^2 + \sigma_j^2)$。因此，

$$\left\|\frac{1}{2}(\mathbf{y}_j^{(1)[1]} + \mathbf{y}_j^{(1)[2]})\right\| = \sqrt{\frac{d_1 - 1}{n}}\sqrt{\sigma_0^2\sigma_j^2 + \sigma_0^4} + O_p\left(\frac{\sqrt{d_1 - 1}}{n}\right) + O_p\left(\frac{1}{\sqrt{n}}\right).$$

(55) 可以用类似的方法证明。

D.2.2 方程(56)。

回顾一下，

$$\begin{aligned}
&\tilde{\Sigma}(\cdot, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]}, \grave{\mathbf{u}}_j^{(1),[1]}, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]}) \\
=&\sum_{l=1}^{r}\sigma_l^2\left(\frac{2}{n}\sum_{k=1}^{n/2}\theta_{lk}^2\right)\left(\prod_{q=2}^{p}\langle\mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)^2\langle\mathbf{u}_l^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle\mathbf{u}_l^{(1)} \\
&+\sum_{l_1\neq l_2}\sigma_{l_1}\sigma_{l_2}\left(\frac{2}{n}\sum_{k=1}^{n/2}\theta_{l_1k}\theta_{l_2k}\right)\left(\prod_{q=2}^{p}\langle\mathbf{u}_{l_1}^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\left(\prod_{q=2}^{p}\langle\mathbf{u}_{l_2}^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\langle\mathbf{u}_{l_2}^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle\mathbf{u}_{l_1}^{(1)} \\
&+\frac{2}{n}\sum_{k=1}^{n/2}\sum_{l=1}^{r}\sigma_l\theta_{lk}\left(\prod_{q=2}^{p}\langle\mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\mathscr{E}_k(\grave{\mathbf{u}}_j^{(1),[1]}, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]})\mathbf{u}_l^{(1)} \\
&+\frac{2}{n}\sum_{k=1}^{n/2}\sum_{l=1}^{r}\mathscr{E}_k(\cdot, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]})\sigma_l\theta_{lk}\left(\prod_{q=2}^{p}\langle\mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\langle\mathbf{u}_l^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle \\
&+\frac{2}{n}\sum_{k=1}^{n/2}\mathscr{E}_k(\cdot, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]})\mathscr{E}_k(\grave{\mathbf{u}}_j^{(1),[1]}, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]}) - \grave{\mathbf{u}}_j^{(1),[1]},
\end{aligned}$$

因此，根据定义，

$$
\begin{aligned}
\mathbf{z}_j^{(1)[1]} = & \sum_{l\neq j, l\in[r]} \sigma_l^2 \left(\frac{2}{n}\sum_{k=1}^{n/2}\theta_{lk}^2\right)\left(\prod_{q=2}^{p}\langle \mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)^2 \langle \mathbf{u}_l^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle \mathbf{u}_l^{(1)} \\
& + \sum_{l_1\neq l_2, l_1\neq j} \sigma_{l_1}\sigma_{l_2}\left(\frac{2}{n}\sum_{k=1}^{n/2}\theta_{l_1k}\theta_{l_2k}\right)\left(\prod_{q=2}^{p}\langle \mathbf{u}_{l_1}^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\left(\prod_{q=2}^{p}\langle \mathbf{u}_{l_2}^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\langle \mathbf{u}_{l_2}^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle \mathbf{u}_{l_1}^{(1)} \\
& + \frac{2}{n}\sum_{k=1}^{n/2}\sum_{l\neq j, l\in[r]} \sigma_l\theta_{lk}\left(\prod_{q=2}^{p}\langle \mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\mathscr{E}_k(\grave{\mathbf{u}}_j^{(1),[1]}, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]})\mathbf{u}_l^{(1)} \\
& + \frac{2}{n}\sum_{k=1}^{n/2}\sum_{l\neq j, l\in[r]} \mathscr{E}_k(\cdot, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]})\sigma_l\theta_{lk}\left(\prod_{q=2}^{p}\langle \mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\langle \mathbf{u}_l^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle \\
& + \mathcal{P}_{\mathbf{u}_j^{(1)}}^{\perp}\left[\frac{2}{n}\sum_{k=1}^{n/2}\mathscr{E}_k(\cdot, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]})\sigma_j\theta_{jk}\left(\prod_{q=2}^{p}\langle \mathbf{u}_j^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\langle \mathbf{u}_j^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle\right. \\
& \qquad \left. - \frac{2}{n}\sum_{k=1}^{n/2}\mathscr{E}_k(\cdot, \mathbf{u}_j^{(2)}, \ldots, \mathbf{u}_j^{(p)})\sigma_j\theta_{jk}\right] \\
& + \mathcal{P}_{\mathbf{u}_j^{(1)}}^{\perp}\left[\frac{2}{n}\sum_{k=1}^{n/2}\mathscr{E}_k(\cdot, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]})\mathscr{E}_k(\grave{\mathbf{u}}_j^{(1),[1]}, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]}) - \grave{\mathbf{u}}_j^{(1),[1]}\right. \\
& \qquad \left. - \frac{2}{n}\sum_{k=1}^{n/2}\mathscr{E}_k(\cdot, \mathbf{u}_j^{(2)}, \ldots, \mathbf{u}_j^{(p)})\mathscr{E}_k(\mathbf{u}_j^{(1)}, \ldots, \mathbf{u}_j^{(p)}) + \mathbf{u}_j^{(1)}\right], \qquad (61)
\end{aligned}
$$

我们将逐一处理这六个项。

第一项。

$$
\mathbf{z}_1^{[1]} := \sum_{l\neq j, l\in[r]} \sigma_l^2 \left(\frac{2}{n}\sum_{k=1}^{n/2}\theta_{lk}^2\right)\left(\prod_{q=2}^{p}\langle \mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)^2 \langle \mathbf{u}_l^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle \mathbf{u}_l^{(1)}.
$$

注意到

$$
\left\|\mathbf{z}_1^{[1]}\right\| \leq \left(\delta^{(1)[2]}\right)^{2p-1} \sum_{l\neq j, l\in[r]} \sigma_l^2 \left|\frac{2}{n}\sum_{k=1}^{n/2}\theta_{lk}^2\right| = O_p\left[\left(\sqrt{\frac{d}{n}}\right)^{2p-1}\right] = O_p\left[\left(\frac{d}{n}\right)^{3/2}\right]. \qquad (62)
$$

最后两个等式由（44）得出，且 $p \geq 2$。由于假设 $d = o(n)$，可推出

$$
\left\|\mathbf{z}_1^{[1]}\right\| \leq O_p\left(\frac{d}{n}\right).
$$

第二项。

$$\begin{aligned}\mathbf{z}_2^{[1]} :=& \sum_{l_1\neq l_2, l_1\neq j} \sigma_{l_1}\sigma_{l_2}\left(\frac{2}{n}\sum_{k=1}^{n/2}\theta_{l_1k}\theta_{l_2k}\right)\left(\prod_{q=2}^{p}\langle \mathbf{u}_{l_1}^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\left(\prod_{q=2}^{p}\langle \mathbf{u}_{l_2}^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\langle \mathbf{u}_{l_2}^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle \mathbf{u}_{l_1}^{(1)} \\ =& \sum_{l_1\neq l_2, l_1\neq j, l_2\neq j} \sigma_{l_1}\sigma_{l_2}\left(\frac{2}{n}\sum_{k=1}^{n/2}\theta_{l_1k}\theta_{l_2k}\right)\left(\prod_{q=2}^{p}\langle \mathbf{u}_{l_1}^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\left(\prod_{q=2}^{p}\langle \mathbf{u}_{l_2}^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\langle \mathbf{u}_{l_2}^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle \mathbf{u}_{l_1}^{(1)} \\ &+ \sum_{l_1\neq j} \sigma_{l_1}\sigma_{j}\left(\frac{2}{n}\sum_{k=1}^{n/2}\theta_{l_1k}\theta_{jk}\right)\left(\prod_{q=2}^{p}\langle \mathbf{u}_{l_1}^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\left(\prod_{q=2}^{p}\langle \mathbf{u}_{j}^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\langle \mathbf{u}_{j}^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle \mathbf{u}_{l_1}^{(1)}.\end{aligned}$$

因为

$$\left|\frac{2}{n}\sum_{k=1}^{n/2}\theta_{l_1k}\theta_{l_2k}\right| = O_p(\frac{1}{\sqrt{n}}),\ \forall l_1\neq l_2,$$

我们立即得到

$$\begin{aligned}&\left\|\mathbf{z}_2^{[1]}\right\| \\ \leq& \left(\delta^{(1)[2]}\right)^{2p-1}\sum_{l_1\neq l_2, l_1\neq j, l_2\neq j}\sigma_{l_1}\sigma_{l_2}\left|\frac{2}{n}\sum_{k=1}^{n/2}\theta_{l_1k}\theta_{l_2k}\right| + \left(\delta^{(1)[2]}\right)^{p-1}\sum_{l_1\neq j}\sigma_{l_1}\sigma_j\left|\frac{2}{n}\sum_{k=1}^{n/2}\theta_{l_1k}\theta_{jk}\right| \\ =& O_p(\frac{1}{\sqrt{n}})\cdot O_p\left[\left(\sqrt{\frac{d}{n}}\right)^{2p-1}\right] + O_p(\frac{1}{\sqrt{n}})\cdot\left[\left(\sqrt{\frac{d}{n}}\right)^{p-1}\right] \\ =& O_p\left(\frac{\sqrt{d}}{n}\right).\end{aligned} \tag{63}$$

第三项。

$$\mathbf{z}_3^{[1]} := \frac{2}{n}\sum_{k=1}^{n/2}\sum_{l\neq j, l\in[r]}\sigma_l\theta_{lk}\left(\prod_{q=2}^{p}\langle \mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\mathscr{E}_k(\grave{\mathbf{u}}_j^{(1),[1]}, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]})\mathbf{u}_l^{(1)}.$$

注意到

$$\begin{aligned}\left\|\mathbf{z}_3^{[1]}\right\| \leq& \sum_{l\neq j, l\in[r]}\sigma_l\left|\prod_{q=2}^{p}\langle \mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right|\cdot\left|\frac{2}{n}\sum_{k=1}^{n/2}\theta_{lk}\mathscr{E}_k(\grave{\mathbf{u}}_j^{(1),[1]}, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]})\right| \\ \leq& r\sigma_1(\delta^{(1)[2]})^{p-1}\cdot\|\widehat{\Sigma}_{\theta,\mathscr{E}}^{[1]}\| \\ =& O_p\left(\sqrt{\frac{d}{n}}\right)^{p-1}\cdot O_p\left(\sqrt{\frac{d}{n}}\right) \\ =& O_p\left(\frac{d}{n}\right),\end{aligned} \tag{64}$$

其中倒数第二个等式由（44）和（43）得出。

第四项。

$$\mathbf{z}_4^{[1]} := \frac{2}{n}\sum_{k=1}^{n/2}\sum_{l\neq j, l\in[r]} \mathscr{E}_k(\cdot, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]})\sigma_l\theta_{lk}\left(\prod_{q=2}^{p}\langle \mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\langle \mathbf{u}_l^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle.$$

类似于 $\mathbf{z}_3^{[1]}$,

$$\left\|\mathbf{z}_4^{[1]}\right\| \leq \|\widehat{\Sigma}_{\theta,\mathscr{E}}\| \cdot \left(\delta^{(1)[2]}\right)^p = O_p\left(\frac{d^{3/2}}{n^{3/2}}\right) = O_p\left(\frac{d}{n}\right).$$

第五项。

$$\begin{aligned}\mathbf{z}_5^{[1]} :=& \mathcal{P}_{\mathbf{u}_j^{(1)}}^{\perp}\left[\frac{2}{n}\sum_{k=1}^{n/2}\mathscr{E}_k(\cdot, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]})\sigma_j\theta_{jk}\left(\prod_{q=2}^{p}\langle \mathbf{u}_j^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\langle \mathbf{u}_j^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle\right.\\ &\left. - \frac{2}{n}\sum_{k=1}^{n/2}\mathscr{E}_k(\cdot, \mathbf{u}_j^{(2)}, \ldots, \mathbf{u}_j^{(p)})\sigma_j\theta_{jk}\right].\end{aligned}$$

我们将引入记号 $\uplus$，表示展开所有

$$\widehat{\mathbf{u}}_j^{(q)[2]} = \mathbf{u}_j^{(q)} + \Delta_j^{(q)[2]},$$

and

$$\grave{\mathbf{u}}_j^{(1),[1]} = \mathbf{u}_j^{(1)} + \grave{\Delta}_j^{(1)[1]},$$

然后保留其中至少含有一个 $\Delta_j^{(\cdot)[2]}$ 或 $\grave{\Delta}_j^{(1)[1]}$ 的所有项。例如，在这里，展开

$$\frac{2}{n}\sum_{k=1}^{n/2}\mathscr{E}_k(\cdot, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]})\sigma_j\theta_{jk}\left(\prod_{q=2}^{p}\langle \mathbf{u}_j^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\langle \mathbf{u}_j^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle$$

会得到 2^{2p-1} 项，我们保留除该项以外的所有内容

$$\frac{2}{n}\sum_{k=1}^{n/2}\mathscr{E}_k(\cdot, \mathbf{u}_j^{(2)}, \ldots, \mathbf{u}_j^{(p)})\sigma_j\theta_{jk}\left(\prod_{q=1}^{p}\langle \mathbf{u}_j^{(q)}, \mathbf{u}_j^{(q)}\rangle\right).$$

于是我们有:

$$\mathbf{z}_5^{[1]} = \mathcal{P}_{\mathbf{u}_j^{(1)}}^{\perp}\left\{\biguplus\left[\frac{2}{n}\sum_{k=1}^{n/2}\mathscr{E}_k(\cdot, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]})\sigma_j\theta_{jk}\left(\prod_{q=2}^{p}\langle \mathbf{u}_j^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\langle \mathbf{u}_j^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle\right]\right\}.$$

那么,

$$\left\|\mathbf{z}_5^{[1]}\right\| \leq \left\|\biguplus\left[\frac{2}{n}\sum_{k=1}^{n/2}\mathscr{E}_k(\cdot, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]})\sigma_j\theta_{jk}\left(\prod_{q=1}^{p}\langle \mathbf{u}_j^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\right]\right\|,$$

并且，对于 $\uplus$ 中的每一项，其范数都满足以下界

$$\|\widehat{\Sigma}_{\theta,\mathscr{E}}\| \cdot \delta^{(1)[2]} = O_p\Big(\frac{d}{n}\Big),$$

从而

$$\left\|\mathbf{z}_5^{[1]}\right\| = O_p\left(\frac{d}{n}\right).$$

第六　项。这一项可以用与上一项类似的方法处理　　.

$$\begin{aligned}
\mathbf{z}_6^{[1]} :=& \mathcal{P}_{\mathbf{u}_j^{(1)}}^{\perp}\left[\frac{2}{n}\sum_{k=1}^{n/2}\mathscr{E}_k(\cdot,\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]})\mathscr{E}_k(\grave{\mathbf{u}}_j^{(1),[1]},\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]}) - \grave{\mathbf{u}}_j^{(1),[1]}\right.\\
&\left. - \frac{2}{n}\sum_{k=1}^{n/2}\mathscr{E}_k(\cdot,\mathbf{u}_j^{(2)},\ldots,\mathbf{u}_j^{(p)})\mathscr{E}_k(\mathbf{u}_j^{(1)},\ldots,\mathbf{u}_j^{(p)}) + \mathbf{u}_j^{(1)}\right]\\
=& \mathcal{P}_{\mathbf{u}_j^{(1)}}^{\perp}\left[\left(\frac{2}{n}\sum_{k=1}^{n/2}\mathscr{E}_k\otimes\mathscr{E}_k - \mathscr{I}\right)(\cdot,\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]},\grave{\mathbf{u}}_j^{(1),[1]},\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]})\right.\\
&\left. - \left(\frac{2}{n}\sum_{k=1}^{n/2}\mathscr{E}_k\otimes\mathscr{E}_k - \mathscr{I}\right)(\cdot,\mathbf{u}_j^{(2)},\ldots,\mathbf{u}_j^{(p)},\mathbf{u}_j^{(1)},\ldots,\mathbf{u}_j^{(p)})\right]\\
=& \mathcal{P}_{\mathbf{u}_j^{(1)}}^{\perp}\left\{\biguplus\left[\left(\frac{2}{n}\sum_{k=1}^{n/2}\mathscr{E}_k\otimes\mathscr{E}_k - \mathscr{I}\right)(\cdot,\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]},\grave{\mathbf{u}}_j^{(1),[1]},\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]})\right]\right\},
\end{aligned}$$

于是，

$$\left\|\mathbf{z}_6^{[1]}\right\| \le \left\|\biguplus\left[\left(\frac{2}{n}\sum_{k=1}^{n/2}\mathscr{E}_k\otimes\mathscr{E}_k - \mathscr{I}\right)(\cdot,\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]},\widehat{\mathbf{u}}_j^{(1)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]})\right]\right\|,$$

并且对于 $\uplus$ 中的每一项，其范数均有界于

$$\|\widehat{\Sigma}_{\mathscr{E}} - \mathscr{I}\| \cdot \delta^{(1)[2]} = O_p\Big(\frac{d}{n}\Big),$$

从而

$$\left\|\mathbf{z}_6^{[1]}\right\| = O_p\left(\frac{d}{n}\right).$$

D.2.3 方程(53)

现在我们证明

$$\langle\grave{\mathbf{u}}_j^{(1),[1]},\mathbf{u}_j^{(1)}\rangle = \sigma_j^2 + O_p\left(\frac{d}{n}\right).$$

由对称性可立即得出

$$\langle \grave{\mathbf{u}}_j^{(1),[2]}, \mathbf{u}_j^{(1)} \rangle = \sigma_j^2 + O_p\left(\frac{d}{n}\right).$$

根据定义,

$$\begin{aligned}
&\langle \grave{\mathbf{u}}_j^{(1),[1]}, \mathbf{u}_j^{(1)} \rangle \\
=&\sigma_j^2 \left(\frac{2}{n}\sum_{k=1}^{n/2}\theta_{jk}^2\right)\left(\prod_{q=2}^{p}\langle \mathbf{u}_j^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)^2 \langle \mathbf{u}_j^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle \\
&+\sum_{l\neq j}\sigma_l\left(\frac{2}{n}\sum_{k=1}^{n/2}\theta_{jk}\theta_{lk}\right)\left(\prod_{q=2}^{p}\langle \mathbf{u}_j^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\left(\prod_{q=2}^{p}\langle \mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\langle \mathbf{u}_l^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle \\
&+\frac{2}{n}\sum_{k=1}^{n/2}\sigma_j\theta_{jk}\left(\prod_{q=2}^{p}\langle \mathbf{u}_j^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\mathscr{E}_k(\grave{\mathbf{u}}_j^{(1),[1]}, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]}) \\
&+\frac{2}{n}\sum_{k=1}^{n/2}\sum_{l=1}^{r}\mathscr{E}_k(\mathbf{u}_j^{(1)}, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]})\sigma_l\theta_{lk}\left(\prod_{q=2}^{p}\langle \mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\langle \mathbf{u}_j^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle \\
&+\frac{2}{n}\sum_{k=1}^{n/2}(\mathscr{E}_k\otimes\mathscr{E}_k-\mathscr{I})(\mathbf{u}_j^{(1)}, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]}, \grave{\mathbf{u}}_j^{(1),[1]}, \widehat{\mathbf{u}}_j^{(2)[2]}, \ldots, \widehat{\mathbf{u}}_j^{(p)[2]}). \qquad (65)
\end{aligned}$$

除第一项外，所有项都将采用与我们界定 $\mathbf{z}_j^{(1)[1]}$ 中六个项时类似的技术进行界定，部分细节从略。

第一项。注意到

$$1-\langle \mathbf{u}_j^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle \le \left(\delta^{(1)[2]}\right)^2,\ 1-\langle \mathbf{u}_j^{(1)}, \grave{\mathbf{u}}_j^{(1)[1]}\rangle \le \left(\delta^{(1)[2]}\right)^2.$$

因此,

$$\sigma_j^2 \left(\frac{2}{n}\sum_{k=1}^{n/2}\theta_{jk}^2\right)\left(\prod_{q=2}^{p}\langle \mathbf{u}_j^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)^2 \langle \mathbf{u}_j^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle = \sigma_j^2 + O_p\left(\frac{d}{n}\right).$$

第二项。

$$\begin{aligned}
&\sum_{l\neq j}\sigma_l\left(\frac{2}{n}\sum_{k=1}^{n/2}\theta_{jk}\theta_{lk}\right)\left(\prod_{q=2}^{p}\langle \mathbf{u}_j^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\left(\prod_{q=2}^{p}\langle \mathbf{u}_l^{(q)}, \widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\langle \mathbf{u}_l^{(1)}, \grave{\mathbf{u}}_j^{(1),[1]}\rangle \\
=&O_p(\frac{1}{\sqrt{n}})\cdot O_p\left[\left(\sqrt{\frac{d}{n}}\right)^{2p-1}\right] \\
=&o_p(\frac{1}{\sqrt{n}}).
\end{aligned}$$

第三项。请注意

$$\begin{aligned}
&\frac{2}{n}\sum_{k=1}^{n/2}\sigma_j\theta_{jk}\left(\prod_{q=2}^{p}\langle \mathbf{u}_j^{(q)},\widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\mathscr{E}_k(\grave{\mathbf{u}}_j^{(1),[1]},\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]})\\
=&\frac{2}{n}\sum_{k=1}^{n/2}\sigma_j\theta_{jk}\left(\prod_{q=2}^{p}\langle \mathbf{u}_j^{(q)},\widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\mathscr{E}_k(\mathbf{u}_j^{(1),[1]},\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]})\\
&+\frac{2}{n}\sum_{k=1}^{n/2}\sigma_j\theta_{jk}\left(\prod_{q=2}^{p}\langle \mathbf{u}_j^{(q)},\widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\mathscr{E}_k(\grave{\mathbf{u}}_j^{(1),[1]}-\mathbf{u}_j^{(1),[1]},\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]}),
\end{aligned}$$

其中第一项为 $O_p(1/\sqrt{n})$，因为$\theta_{jk}\mathscr{E}_k$，$k\in[n/2]$ 相互独立，且 $\widehat{\ }\mathbf{u}_j^{(2)[2]}$，…，$\widehat{\ }\mathbf{u}_j^{(p)[2]}$，第二项的上界为

$$\|\widehat{\Sigma}_{\theta,\mathscr{E}}\|\cdot\delta^{(1)[2]}=O_p\Big(\frac{d}{n}\Big),$$

so

$$\frac{2}{n}\sum_{k=1}^{n/2}\sigma_j\theta_{jk}\left(\prod_{q=2}^{p}\langle \mathbf{u}_j^{(q)},\widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\mathscr{E}_k(\grave{\mathbf{u}}_j^{(1),[1]},\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]})=O_p\left(\frac{d}{n}\right)+O_p\left(\frac{1}{\sqrt{n}}\right)=O_p\left(\frac{d}{n}\right).$$

最后一个等式由假设 $d=o(n)$ 得出。

第四项。回想一下，$\widehat{\ }\mathbf{u}_j^{(q)[2]}$ 是数据后半部分的函数，它与数据前半部分独立，即与所有索引为 $k\le n/2$ 的随机变量独立。因此，条件于数据后半部分，对于任意 $l\in[r]$,

$$\mathscr{E}_k(\mathbf{u}_j^{(1)},\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]})\theta_{lk}$$

是两个相互独立的 $N(0,\ \sigma_0^2)$ 和 $N(0,\ 1)$ 随机变量的乘积，并且它们在 $k\le n/2$ 上独立同分布。因此我们有

$$\frac{2}{n}\sum_{k=1}^{n/2}\sum_{l=1}^{r}\mathscr{E}_k(\mathbf{u}_j^{(1)},\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]})\sigma_l\theta_{lk}\left(\prod_{q=2}^{p}\langle \mathbf{u}_l^{(q)},\widehat{\mathbf{u}}_j^{(q)[2]}\rangle\right)\langle \mathbf{u}_j^{(1)},\grave{\mathbf{u}}_j^{(1),[1]}\rangle=O_p(\frac{1}{\sqrt{n}}).$$

第五项。与 $\mathbf{z}_j^{(1)[1]}$ 中的第六项类似,

$$\frac{2}{n}\sum_{k=1}^{n/2}(\mathscr{E}_k\otimes\mathscr{E}_k-\mathscr{I})(\mathbf{u}_j^{(1)},\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]},\grave{\mathbf{u}}_j^{(1),[1]},\widehat{\mathbf{u}}_j^{(2)[2]},\ldots,\widehat{\mathbf{u}}_j^{(p)[2]}).$$

综合以上五个上界，我们得到

$$\langle\grave{\mathbf{u}}_j^{(1),[1]},\mathbf{u}_j^{(1)}\rangle=\sigma_j^2+O_p\left(\frac{d}{n}\right)+O_p\left(\frac{1}{\sqrt{n}}\right)=O_p\left(\frac{d}{n}\right).$$

最后一个等式由假设 $d=o(n)$ 得出。

D.3 步骤 3。

回忆一下,

$$\widehat{b}_j^{(q)[1]} := \left\langle \widehat{\mathbf{u}}_j^{(q),[1][1]}, \widehat{\mathbf{u}}_j^{(q),[1][2]} \right\rangle - 1,$$
$$\widehat{b}_j^{(q)[2]} := \left\langle \widehat{\mathbf{u}}_j^{(q),[2][1]}, \widehat{\mathbf{u}}_j^{(q),[2][2]} \right\rangle - 1.$$

在这一部分，我们使用引理8来证明

$$\sqrt{n}\left(\widehat{b}_j^{(1)[1]} - b_j^{(1)[1]}\right) \xrightarrow{p} 0, \tag{66}$$

其中 $b_j^{(1)[1]}$ 定义见引理 3。类似地 $\sqrt{n}\left(\widehat{b}_j^{(1)[2]} - b_j^{(1)[2]}\right) \xrightarrow{p} 0$。然后，回顾根据定义（39），

$$b_j^{(1)} = \frac{\left\|\check{\mathbf{u}}_j^{(1),[1]} + \check{\mathbf{u}}_j^{(1),[2]}\right\|}{\sqrt{1+b_j^{(1)[1]}} + \sqrt{1+b_j^{(1)[2]}}} - 1,$$

根据定义（13），

$$\widehat{b}_j^{(1)} = \frac{\left\|\check{\mathbf{u}}_j^{(q),[1]} + \check{\mathbf{u}}_j^{(q),[2]}\right\|}{\sqrt{1+\widehat{b}_j^{(1)[1]}} + \sqrt{1+\widehat{b}_j^{(1)[2]}}} - 1,$$

我们有（42）可由（66）以及以下事实推出： $\left\|\check{\mathbf{u}}_j^{(1)\,[1]} + \check{\mathbf{u}}_j^{(1)\,[2]}\right\| \leq 2$。现在我们来关注（66）。

根据引理 4,

$$\begin{aligned}
&\bar{\mathbb{E}}\|\widehat{M}_j^{[1]} - M_j^{[1]}\| \\
\leq & C\|M_j^{[1]}\|\left(\sqrt{\frac{r(M_j^{[1]})}{n}} \bigvee \frac{r(M_j^{[1]})}{n}\right) \\
\leq & \frac{C(\sigma_j^2+1)(\sigma_j^2+d_1+1)}{(\frac{1}{2}\sigma_j^2+1)n},
\end{aligned}$$

在事件 $\mathcal{B}_n$. $(M_j^{[1]}, \widehat{M}_j^{[1]}$ 以及事件 $\mathcal{B}_n$ 在引理3的证明开头处定义。）最后一个不平等由（98）和（94）成立。请记住，$\bar{g}_1 \geq \frac{1}{2}\sigma_j^2$（不平等（96））, 因此在假设 $d = o(n$的条件下，对于足够大的 n， $\bar{\mathbb{E}}\|\widehat{M}_j^{[1]} - M_j^{[1]}\| \leq \frac{1}{4}\bar{g}_1$，也就是说，若令 $\gamma = \frac{1}{4}$，则

$$\bar{\mathbb{E}}\|\widehat{M}_j^{[1]} - M_j^{[1]}\| \leq \frac{(1-2\gamma)\bar{g}_1}{2}.$$

$1+b_j^{(1)[1]} \geq 2\gamma$ 对于足够大的 n，由不平等（99）也可得出。令 $t = \left(\begin{smallmatrix} n(M_j^{[1]}) \\ 2r \end{smallmatrix}\right)$。首先注意，在事件 $\mathcal{B}_n$ 上，

$$t \geq \left(\frac{n(\frac{1}{2}\sigma_j^2+1)}{2(\sigma_j^2+d_1+1)}\right)^{\frac{1}{3}},$$

由 in 等式 (98)，假设 $d = o(n)$，我们有 $t \geq 1$ 对于足够大的 n .

令 $D := D_1$ 如引理 8中所示。我们有

$$D\|M_j^{[1]}\|\left(\sqrt{\frac{t}{n}} \bigvee \sqrt{\frac{\log n}{n}}\right) \leq D(\sigma_j^2+1)\left(n^{-\frac{2}{3}} \bigvee \sqrt{\frac{\log n}{n}}\right) \leq \frac{1}{8}\bar{g}_1$$

对于足够大的 n。

因此，引理 8中的所有条件均满足，其中 $\widehat{M}_j^{[1]}$和 $M_j^{[1]}$，条件于$\{\mathscr{X}_k,\ k = n/2+1, \ldots\ldots, n\}$ 在事件 $\mathcal{B}_n$下，对于足够大的 n。

注意，条件于 $\{\mathscr{X}_k,\ k = n/2+1, \ldots\ldots, n\}$，$\widehat{b}_j^{(1)[1]}$和 $b_j^{(1)[1]}$正是定义为 $\widehat{b}_1$ 和 b_1 [在 (78) 和 (79)] 中对应于 $M_j^{[1]}$ 的定义。现在我们可以应用引理8:

$$\begin{aligned}
&\sqrt{n}\left|\widehat{b}_j^{(1)[1]} - b_j^{(1)[1]}\right| \\
\leq& D\sqrt{n}\frac{\|M_j^{[1]}\|^2}{\bar{g}_1^2}\left(\sqrt{\frac{r(M_j^{[1]})}{n}} \bigvee \sqrt{\frac{t}{n}} \bigvee \sqrt{\frac{\log n}{n}}\right)\sqrt{\frac{t}{n}} \\
\leq& D\frac{\|M_j^{[1]}\|^2}{\bar{g}_1^2}\left(\left(\frac{r(M_j^{[1]})}{n}\right)^{\frac{1}{3}} \bigvee \frac{1}{r(M_j^{[1]})^{\frac{1}{3}}n^{\frac{1}{6}}} \bigvee \frac{\sqrt{\log n}}{r(M_j^{[1]})^{\frac{1}{6}}n^{\frac{1}{3}}}\right) \\
\leq& D\frac{\|M_j^{[1]}\|^2}{\bar{g}_1^2}\left(\left(\frac{r(M_j^{[1]})}{n}\right)^{\frac{1}{3}} \bigvee \frac{1}{n^{\frac{1}{6}}}\right) \\
\leq& D\left(\frac{\sigma_j^2+1}{\frac{1}{2}\sigma_j^2}\right)^2\left[\left(\frac{\sigma_j^2+d_1+1}{n(\frac{1}{2}\sigma_j^2+1)}\right)^{\frac{1}{3}} \bigvee \frac{1}{n^{\frac{1}{6}}}\right]
\end{aligned} \tag{67}$$

条件于$\{\mathscr{X}_k,\ k = n/2+1, \ldots\ldots, n\}$ 在事件$\mathcal{B}_n$下，当 n足够大时，概率至少为 $1-e^{-t}$。最后一个不平等成立，是因为（94）、（96）和（98）。

由于在事件 $\mathcal{B}_n$上，条件于 $\{\mathscr{X}_{n/2+1}, \ldots, \mathscr{X}_n\}$，(67) 成立的概率至少为 $1-e^{-t} \geq 1-\exp\left[-\left(\begin{smallmatrix} n \\ 2 \end{smallmatrix}\right)^{\frac{3}{3}}_{(\frac{1}{2}\sigma_j^2+1)(\sigma_j^2+d_1+1)}\right]$，因此在事件 $\mathcal{B}_n$上，(67) 也以如下概率成立

概率至少为 $1-\exp\left[-\left(\frac{n}{2^{(\frac{1}{2}\sigma_j^2+1)}(\sigma_j^2+d_1+1)}\right)^{\frac{1}{3}}\right]$。结合假设 $\frac{d}{n}\to 0$，可得

$$\sqrt{n}\left(\widehat{b}_j^{(1)[1]}-b_j^{(1)[1]}\right)\cdot\mathbb{I}(\mathcal{B}_n)\xrightarrow{p}0,$$

由此可得

$$\sqrt{n}\left(\widehat{b}_j^{(1)[1]}-b_j^{(1)[1]}\right)\xrightarrow{p}0.$$

E 引理证明3

引理 3 的证明大量依赖于 Koltchinskii 和 Lounici (2014) 的技巧和结果，我们将在此先行回顾。

E.1 预备知识

设 $\mathbb{H}$ 是一个希尔伯特空间，$M:\ \mathbb{H}\to\mathbb{H}$ 是一个紧对称半正定算子。众所周知，以下谱表示成立

$$M=\sum_{r>1}\mu_r\mathcal{P}_r$$

其中不同的非零特征值 μ_r 按降序排列为 $\mu_1>\mu_2>\cdots\geq 0$，且$\mathcal{P}_r$ 是相应的谱投影算子。M 的有效秩定义为

$$r(M):=\frac{\mathrm{tr}(M)}{\|M\|}.$$

我们将特别使用Koltchinskii和Lounici（2014）中关于估计$\mathcal{P}_1$,的结果，适用于

$$\mathcal{P}_1=\mathbf{u}_1\otimes\mathbf{u}_1,$$

即，在主特征值为孤立单特征值的情况下估计主特征向量。令 Y_1，Y_2，……，Y_n 为 $\mathbb{R}^m$ 中具有协方差 $M=\mathbb{E}(Y\otimes Y)$ 的独立同分布的中心化高斯随机向量。令

$$\widehat{M}:=\frac{1}{n}\sum_{k=1}^{n}Y_k\otimes Y_k$$

是基于以下观测（Y_1、Y_2、……、Y_n）得到的样本协方差矩阵。以下引理是对 Koltchinskii 和 Lounici （2014）中定理 1 的重新表述。

引理 4。

$$\mathbb{E}\|\widehat{M}-M\| \asymp \|M\|\left(\sqrt{\frac{r(M)}{n}}\bigvee\frac{r(M)}{n}\right),$$

and

$$\mathbb{E}\|\widehat{M}-M\|^2 \asymp \|M\|^2\left(\sqrt{\frac{r(M)}{n}}\bigvee\frac{r(M)}{n}\right)^2.$$

令 $\hat{\mathbf{u}}_1$ 为 M 的主特征向量。为使 $\hat{\mathbf{u}}_1$ 的线性形式定义良好，不失一般性，我们总是假设 $\langle\hat{\mathbf{u}}_1,\mathbf{u}_1\rangle \geq 0$。记 $\mathcal{P}_1 := \hat{\mathbf{u}}_1 \otimes \hat{\mathbf{u}}_1$。定义$\bar{g}_1 := \mu_1 - \mu_2$，即 μ_1 的谱隙，并记

$$\mathcal{C}_1 = \sum_{s\neq 1}\frac{1}{\mu_1-\mu_s}\mathcal{P}_s, \tag{68}$$

$$\mathcal{L}_1 := \mathcal{C}_1(\widehat{M}-M)\mathcal{P}_1 + \mathcal{P}_1(\widehat{M}-M)\mathcal{C}_1 = \frac{1}{n}\sum_{j=1}^{n}(\mathcal{C}_1Y_j\otimes\mathcal{P}_1Y_j + \mathcal{P}_1Y_j\otimes\mathcal{C}_1Y_j), \tag{69}$$

$$\mathcal{S}_1 := \widehat{\mathcal{P}}_1 - \mathcal{P}_1 - \mathcal{L}_1, \tag{70}$$

以及用算子表示的余项

$$\mathcal{R}_1 := \widehat{\mathcal{P}}_1 - \mathbb{E}\widehat{\mathcal{P}}_1 - \mathcal{L}_1. \tag{71}$$

注意 $\mathbb{E}\mathcal{L}_1 = 0$，因此 $\mathcal{R}_1 = \mathcal{S}_1 - \mathbb{E}\mathcal{S}_1$。如定理 4.4 的证明中一样，我们用花体大写字母表示 $\mathcal{P}_1$, $\mathcal{P}_1$, $\mathcal{C}_1$, $\mathcal{L}_1$, $\mathcal{S}_1$ 和 $\mathcal{R}_1$ ，以表明它们是矩阵。将 Koltchinskii 和 Lounici (2014) 中的引理 1 改述如下，我们有

引理 5。

$$\|\mathcal{S}_1\| \leq 14\left(\frac{\|\widehat{M}-M\|}{\bar{g}_1}\right)^2 \tag{72}$$

结合引理5 和4，可得

$$\mathbb{E}\|\mathcal{S}_1\| \leq C\frac{\|M\|^2}{\bar{g}_1^2}\left(\sqrt{\frac{r(M)}{n}}\bigvee\frac{r(M)}{n}\right)^2, \tag{73}$$

其中 C 是一个普适常数。重述 Koltchinskii 和 Lounici(2014) 的定理 3 和定理 4，可得

引理 6。 设 $t>1$，并假设存在某个 $\gamma\in(0,1)$ 以及一个充分大的常数$C>0$,

$$\mathbb{E}\|\widehat{M}-M\|+C\|M\|\sqrt{\frac{t}{n}}\leq\frac{1-\gamma}{1+\gamma}\frac{\bar{g}_1}{2}. \tag{74}$$

那么存在一个常数 $D_\gamma>0$，使得对于所有 u、$v\in\mathbb{H}$，以下界以至少 $1-e^{-t}$ 的概率成立:

$$\left|\left\langle(\widehat{\mathcal{P}}_1-\mathbb{E}\widehat{\mathcal{P}}_1-\mathcal{L}_1)u,v\right\rangle\right|\leq D_\gamma\frac{\|M\|^2}{\bar{g}_1^2}\left(\sqrt{\frac{r(M)}{n}}\bigvee\sqrt{\frac{t}{n}}\right)\sqrt{\frac{t}{n}}\|u\|\|v\|. \tag{75}$$

引理 7. 假设对于某个 $\gamma\in(0,\ 1)$ 以及一个充分大的常数 $C>0$,

$$\mathbb{E}\|\widehat{M}-M\|+C\|M\|\frac{\log n}{n}\leq(1-\gamma)\frac{\bar{g}_1}{2}. \tag{76}$$

那么，存在一个常数 $D_\gamma>0$，使得

$$\|\mathbb{E}\widehat{\mathcal{P}}_1-\mathcal{P}_1-\mathcal{P}_1\mathbb{E}(\mathcal{S}_1)\mathcal{P}_1\|\leq D_\gamma\frac{\|M\|^2}{\bar{g}_1^2}\frac{1}{\sqrt{n}}\left(\sqrt{\frac{r(M)}{n}}\bigvee\sqrt{\frac{\log n}{n}}\right). \tag{77}$$

^定义 $\mathcal{S}_1$ 后，我们可以定义一个刻画$\mathcal{P}_1$偏差的临界量:

$$b_1:=\langle\mathbb{E}(\mathcal{S}_1)\mathbf{u}_1,\mathbf{u}_1\rangle. \tag{78}$$

^注意， $\mathbb{E}\mathcal{L}_1=0$ 和$\mathcal{P}_1:=\hat{}\mathbf{u}_1\otimes\hat{}\mathbf{u}_1,$，其中$\langle\hat{}\mathbf{u}_1,\mathbf{u}_1\rangle\geq0$，我们有

$$b_1=\langle\mathbb{E}(\hat{\mathbf{u}}_1\otimes\hat{\mathbf{u}}_1)\mathbf{u}_1,\mathbf{u}_1\rangle-1,$$

因此 $-1\leq b_1\leq0$。

有一种估计 b_1 的方法。假设我们将样本（Y_1，Y_2，.....，Y_n）分成两个样本量均为 $\lfloor\frac{n}{2}\rfloor$ 的子样本。令 $\check{}M$ 为基于第一个 $\check{}$ 子样本得到的样本协方差，M' 为基于第二个子样本得到的样本协方差。记 $\check{}\mathbf{u}_1\check{}\check{}$ 为 M 的主特征向量，$\check{}\mathbf{u}_1'$ 为 M' 的主特征向量。假设选择它们的符号，使得 $\langle\check{}\mathbf{u}_1,\check{}\mathbf{u}_1'\rangle\geq0$。定义

$$\widehat{b}_1:=\langle\check{\mathbf{u}}_1,\check{\mathbf{u}}_1'\rangle-1. \tag{79}$$

下面的引理重述了 Koltchinskii 和 Lounici (2014) 中的命题 3，^给出了 $|b_1-b_1|$ 的集中不等式。

引理 8。 令 $t \geq 1$ 和 $\gamma \in (0,\ 1/2)$。存在一个常数 $D_\gamma > 0$，使得若

$$\mathbb{E}\|\widehat{M} - M\| \leq \frac{(1-2\gamma)\bar{g}_1}{2},\ 1 + b_1 \geq 2\gamma \tag{80}$$

and

$$D_\gamma \|M\| \left(\sqrt{\frac{t}{n}} \bigvee \sqrt{\frac{\log n}{n}}\right) \leq \frac{\gamma \bar{g}_1}{2}, \tag{81}$$

那么，以至少 $1 - e^{-t}$ 的概率，

$$|\widehat{b}_1 - b_1| \leq D_\gamma \frac{\|M\|^2}{\bar{g}_1^2} \left(\sqrt{\frac{r(M)}{n}} \bigvee \sqrt{\frac{t}{n}} \bigvee \sqrt{\frac{\log n}{n}}\right) \sqrt{\frac{t}{n}}. \tag{82}$$

证明思路。 不失一般性，在整个证明过程中，我们假设 $\|\mathbf{u}\|,\|\mathbf{v}\| = 1$。记

$$\mathcal{L}_j := \frac{2}{n} \sum_{k=1}^{n/2} \left(\mathcal{C}_j y_{jk} \otimes \mathcal{P}_j y_{jk} + \mathcal{P}_j y_{jk} \otimes \mathcal{C}_j y_{jk}\right), \tag{83}$$

then

$$\langle \mathcal{L}_j \mathbf{u}, \mathbf{v} \rangle = \frac{2}{n} \sum_{k=1}^{n/2} \left[\langle y_{jk}, \mathcal{P}_j \mathbf{v} \rangle \langle y_{jk}, \mathcal{C}_j \mathbf{u} \rangle + \langle y_{jk}, \mathcal{P}_j \mathbf{u} \rangle \langle y_{jk}, \mathcal{C}_j \mathbf{v} \rangle\right]. \tag{84}$$

现在，引理 3 等价于：存在普适常数 C 使得

$$\mathbb{P}(|b_j^{(q)[1]}| \leq C\frac{d}{n}) \to 1, \tag{85}$$

并且对于 $\forall \mathbf{u},\mathbf{v} \in \mathbb{H},\ \|\mathbf{u}\|, \|\mathbf{v}\| \leq 1$,

$$\sqrt{\frac{n}{2}} \left\langle \left[\widehat{\mathcal{P}}_j - \mathcal{P}_j - \mathcal{P}_j \bar{\mathbb{E}}(\mathcal{S}^{[1]}) \mathcal{P}_j - \mathcal{L}_j\right] \mathbf{u}, \mathbf{v} \right\rangle \xrightarrow{p} 0. \tag{86}$$

我们将（86）的证明分为三部分：

$$\sqrt{\frac{n}{2}} \left\langle (\widehat{\mathcal{P}}_j - \bar{\mathbb{E}}\widehat{\mathcal{P}}_j - \mathcal{L}^{[1]}) u, v \right\rangle \xrightarrow{p} 0, \tag{87}$$

$$\sqrt{\frac{n}{2}} \left\| \bar{\mathbb{E}}\widehat{\mathcal{P}}_j - \mathcal{P}_j - \mathcal{P}_j \bar{\mathbb{E}}(S^{[1]}) \mathcal{P}_j \right\| \xrightarrow{p} 0, \tag{88}$$

$$\sqrt{\frac{n}{2}} \left\langle (\mathcal{L}^{[1]} - \mathcal{L}_j) u, v \right\rangle \xrightarrow{p} 0. \tag{89}$$

我们将首先证明一些初步界以及（85），然后回到（87）和（89）以完成证明。 (88)

E.2 证明（85）。

上界（16）和（17）意味着 $\hat{\mathbf{u}}_j^{(q)[2]}$ 满足以下条件：

$$\lim_{n\to\infty}\mathbb{P}\left(\|\hat{\mathbf{u}}_j^{(q)[2]}-\mathbf{u}_j^{(q)}\|\le a_n,\ \forall q\in[p]\right)=1, \tag{90}$$

其中，a_n 是一个数值序列，使得$\lim_{n\to\infty}a_n=0$，并且

$$\lim_{n\to\infty}\mathbb{P}\left(\max_{l\neq j}\left|\sigma_l^2\langle\mathbf{u}_l^{(q)},\hat{\mathbf{u}}_j^{(q)[2]}\rangle\right|\le A,\ \forall q\in[p]\right)=1, \tag{91}$$

其中，A 是一个数值常数。定义事件

$$\mathcal{B}_n:=\left\{\|\hat{\mathbf{u}}_j^{(q)[2]}-\mathbf{u}_j^{(q)}\|\le a_n,\ \|\hat{\mathbf{u}}_j^{(q)[2]}-\mathbf{u}_j^{(q)}\|\le a_n,\ \forall q\in[p]\right\}.$$

根据（90）和（91），$\lim_{n\to\infty}\mathbb{P}(\mathrm{B}_n)=1$。注意，事件$\mathrm{B}_n$ 属于$\bar{\ }$ $\bar{\ }$的σ-域{$\mathscr{X}_{n/2+1}$, $\mathscr{X}_{n/2+2},\ldots,\mathscr{X}_n$}，因此只需将 $\mathbb{P}(\cdot)$和 $\mathbb{E}(\cdot)$视为条件于事件B_n。由于事件B_n 满足$\lim_{n\to\infty}\mathbb{P}(\mathrm{B}_n)=1$，对于任意随机变量序列 Z_n，要证明 $Z_n\xrightarrow{p}0$，我们只需证明 $Z_n\cdot\mathbb{I}(\mathrm{B}_n)\xrightarrow{p}0$，其中 $\mathbb{I}(\mathrm{B}_n)$是事件B_n的示性函数。我们将广泛使用这一技巧。

我们只需证明 $q=1$的情形。从现在起直到本证明结束，为简化记号，我们记$\mathcal{P}_j=\mathcal{P}_j^{(1)}$、$\mathcal{C}_j=\mathcal{C}_j^{(1)}$、$\hat{\mathcal{P}}_j=\hat{\mathcal{P}}_j^{(1)}$、$M_j^{[1]}=M_{(1)[1]}$,$\mathcal{C}_j^{[1]}=\mathcal{C}_j^{(1)[1]}$、$\mathcal{L}_j^{[1]}=\mathcal{L}_j^{(1)[1]}$、$\mathcal{S}_j^{[1]}=\mathcal{S}_j^{(1)[1]}$。此外，记

$$\mathcal{P}_-=I_{d_q}-\sum_{l=1}^{r}\mathbf{u}_l^{(q)}\otimes\mathbf{u}_l^{(q)} \tag{92}$$

注意，$\mathrm{C}_j^{[1]}$、$\mathrm{L}_j^{[1]}$、$\mathrm{S}_j^{[1]}$, $R^{[1]}$是与我们的协方差矩阵 $M_j^{[1]}$ 相对应、在 (68)、(69)、(70)、(71)] 中定义的 $\mathrm{C}_1,\mathrm{L}_1$,$\mathrm{S}_1$,$\mathrm{R}_1$ [。条件于 {$\mathscr{X}_k, k=n/2+1,\ldots,n$},

$$z_{jk}:=\mathscr{X}_k\times_2\hat{\mathbf{u}}_j^{(1)[2]}\cdots\times_p\hat{\mathbf{u}}_j^{(p)[2]}, k=1,\ldots,n/2$$

其协方差矩阵为 $M_j^{[1]}$，而 $\hat{\mathbf{u}}_j^{(q)[1]}$ 是样本协方差矩阵的主特征向量

$$\widehat{M}_j^{[1]}=\frac{2}{n}\sum_{k=1}^{n/2}z_{jk}\otimes z_{jk}. \tag{93}$$

我们首先在事件 $\mathcal{B}_n$ 下，证明关于$\|M_j^{[1]}\|$、$\bar{g}_1$ $M_j^{[1]}$ 的首个谱隙以及 $r(M_j^{[1]})$（$M_j^{[1]}$ 的有效秩）的一些不等式。这些不等式将在整个证明过程中得到广泛应用。

在 $\mathcal{B}_n$ 下，由于 $\|\hat{\mathbf{u}}_j^{(q)[2]} - \mathbf{u}_j^{(q)}\| \le a_n$，我们有 $(1-a_n^2)^{\frac{p}{2}}\sigma_j^2 \le \tilde{\sigma}_j^2 \le \sigma_j^2$。此外，因为 $\max_{l\neq j}\left|\sigma_l^2 \mathbf{u}_l^{(2)}, \hat{\mathbf{u}}_j^{(2)[2]}\right| \le A$ 有界，且 $\left|\mathbf{u}_l^{(2)}, \hat{\mathbf{u}}_j^{(2)[2]}\right| \le \sqrt{1-\langle \hat{\mathbf{u}}_j^{(2)[2]}, \mathbf{u}_j^{(2)}\rangle^2} \le a_n$，我们有 $\max_{l\neq j}\{\tilde{\sigma}_l^2\} \le Aa_n$。由于 $\lim_{n\to\infty} a_n = 0$，当 n 足够大时，我们有 $(1-a_n^2)^{\frac{p}{2}}\sigma_j^2 > Aa_n$。因此，$M_j^{[1]}$ 的主特征向量为 $\mathbf{u}_j^{(1)}$，对应的特征值为 $\tilde{\sigma}_j^2 + 1$。因此，当 n 足够大时，

$$\frac{1}{2}\sigma_j^2 + 1 \le (1-a_n^2)^{\frac{p}{2}}\sigma_j^2 + 1 \le \|M_j^{[1]}\| \le \sigma_j^2 + 1, \tag{94}$$

以及 $M_j^{[1]}$ 的主特征值的谱隙

$$\bar{g}_1 \ge (1-a_n^2)^{\frac{p}{2}}\sigma_j^2 - Aa_n. \tag{95}$$

由于 $\lim_{n\to\infty} a_n = 0$，当 n 足够大时，我们有：

$$\bar{g}_1 \ge \frac{1}{2}\sigma_j^2. \tag{96}$$

根据定义，

$$r(M_j^{[1]}) = \frac{\mathrm{tr}(M_j^{[1]})}{\|M_j^{[1]}\|}. \tag{97}$$

结合（94），可得

$$\begin{aligned}
&\mathrm{tr}(M_j^{[1]}) \\
=&\sum_{l\neq j}\sigma_l^2\left(\prod_{q=2}^{p}\langle \mathbf{u}_l^{(q)}, \hat{\mathbf{u}}_j^{(q)[2]}\rangle\right)^2 + \tilde{\sigma}_j^2 + d_1 \\
\le&\max_{l\neq j}\left|\sigma_l^2\langle \mathbf{u}_l^{(2)}, \hat{\mathbf{u}}_j^{(2)[2]}\rangle\right| + \sigma_j^2 + d_1 \\
\le&Aa_n + \sigma_j^2 + d_1
\end{aligned}$$

其中，第一个不等式由柯西不等式得出。因此，当 n 足够大时，

$$\frac{d_1}{\sigma_j^2+1} \le r(M_j^{[1]}) \le \frac{\sigma_j^2 + d_1 + 1}{\frac{1}{2}\sigma_j^2 + 1}. \tag{98}$$

结合引理 6 中的不等式（73）与界（94）、（96）及（98），可知在事件 $\mathcal{B}_n$ 下，当 n 足够大时，

$$\left|b_j^{(1)[1]}\right| \le \left\|\bar{\mathbb{E}}(\mathcal{S}_j^{[1]})\right\| \le \bar{\mathbb{E}}\left\|\mathcal{S}_j^{[1]}\right\| \le C\frac{(\sigma_j^2+1)^2}{\frac{1}{4}\sigma_j^4}\left(\sqrt{\frac{\sigma_j^2+d_1+1}{n(\frac{1}{2}\sigma_j^2+1)}} \bigvee \frac{\sigma_j^2+d_1+1}{n(\frac{1}{2}\sigma_j^2+1)}\right)^2, \tag{99}$$

由于我们将 σ_j 视为固定值，右侧有界于 $C\left(\sqrt{\frac{d}{n}} \bigvee \frac{d}{n}\right)^2 \le C\frac{d}{n}$。回想 $\lim_{n\to\infty}\mathbb{P}(\mathcal{B}_n) = 1$，我们有 $\mathbb{P}(|b_j^{(1)[1]}| \le C\frac{d}{n}) \to 1$，当 $n \to \infty$ 时。

E.3 的证明 (87)。

在这一部分，我们将使用引理6 来证明

$$\sqrt{\frac{n}{2}}\left\langle(\widehat{\mathcal{P}}_j - \bar{\mathbb{E}}\widehat{\mathcal{P}}_j - \mathcal{L}_j^{[1]})\mathbf{u}, \mathbf{v}\right\rangle \xrightarrow{p} 0.$$

令 $t = \left(\frac{n(M_j^{[1]})}{2r}\right)^{\frac{1}{3}}$。首先注意，在事件 $\mathcal{B}_n$ 上，由不平等式（98），

$$t \geq \left(\frac{n(\frac{1}{2}\sigma_j^2 + 1)}{2(\sigma_j^2 + d_1 + 1)}\right)^{\frac{1}{3}} \geq 1,$$

当 n 足够大时。由引理 4,

$$\begin{aligned}
\delta_n(t) &= \bar{\mathbb{E}}\|\widehat{M}_j^{[1]} - M_j^{[1]}\| + C\|M_j^{[1]}\|\sqrt{\frac{t}{n}} \\
&\leq C\|M_j^{[1]}\|\left(\sqrt{\frac{r(M_j^{[1]})}{n}} \bigvee \frac{r(M_j^{[1]})}{n}\right) + C\|M_j^{[1]}\|\sqrt{\frac{t}{n}} \\
&\leq C\|M_j^{[1]}\|\left[\frac{4(\sigma_j^2 + d_1 + 1)}{(\frac{1}{2}\sigma_j^2 + 1)n} + \sqrt{\frac{2\left(\frac{n}{2r(M_j^{[1]})}\right)^{\frac{1}{3}}}{n}}\right] \\
&\leq C(\sigma_j^2 + 1)\left[\frac{\sigma_j^2 + d_1 + 1}{(\frac{1}{2}\sigma_j^2 + 1)n} + n^{-\frac{1}{3}}\right],
\end{aligned}$$

在事件 $\mathcal{B}_n$ 上。最后一个不平等式成立是因为（98）和（94）。回想一下，$\bar{g}_1 \geq \frac{1}{2}\sigma_j^2$。我们有，当 n 足够大时，$\delta_n(t) \leq \frac{1}{6}\bar{g}_1$。

现在，根据引理 6及 $t = \left(\frac{n(M_j^{[1]})}{2r}\right)^{\frac{1}{3}}$和 $\gamma = \frac{1}{2}$，存在一个常数 $D := D_{\frac{1}{2}}$，使得在事件 $\mathcal{B}_n$ 上，条件于 $\{\mathscr{X}_{n/2+1}, \mathscr{X}_{n/2+2}, \ldots, \mathscr{X}_n\}$，其概率至少为

least $1-e^{-t} \geq 1-\exp\left[-\left(\frac{n(\frac{1}{2}\sigma_j^2+1)}{2(\sigma_j^2+d_1+1)}\right)^{\frac{1}{3}}\right]$:

$$
\begin{aligned}
&\sqrt{\frac{n}{2}}\left|\left\langle(\widehat{\mathcal{P}}_j-\bar{\mathbb{E}}\widehat{\mathcal{P}}_j-\mathcal{L}_j^{[1]})\mathbf{u},\mathbf{v}\right\rangle\right| \\
\leq & D\sqrt{\frac{n}{2}}\frac{\|M_j^{[1]}\|^2}{\bar{g}_1^2}\left(\sqrt{\frac{r(M_j^{[1]})}{n}}\bigvee\sqrt{\frac{t}{n}}\right)\sqrt{\frac{t}{n}} \\
= & D\frac{\|M_j^{[1]}\|^2}{\bar{g}_1^2}\left[\left(\frac{r(M_j^{[1]})}{n}\right)^{\frac{1}{3}}\bigvee\frac{1}{r(M_j^{[1]})^{\frac{1}{3}}(n/2)^{\frac{1}{6}}}\right] \\
\leq & 4D\left(\frac{\sigma_j^2+1}{\frac{1}{2}\sigma_j^2}\right)^2\left[\left(\frac{\sigma_j^2+d_1+1}{n(\frac{1}{2}\sigma_j^2+1)}\right)^{\frac{1}{3}}\bigvee\frac{1}{n^{\frac{1}{6}}}\right],
\end{aligned}
\tag{100}
$$

条件于 $\{\mathscr{X}_k,\ k=n/2+1,\ \ldots,\ n\}$ 在事件 $\mathcal{B}_n$ 下，对足够大的 n，其概率至少为 $1-e^{-t}$。最后一个不等式由 (94)、(96) 和 (98) 得出。

由于在事件 $\mathcal{B}_n$ 上，条件于 $\{\mathscr{X}_{n/2+1},\mathscr{X}_{n/2+2},\ldots,\mathscr{X}_n\}$ 时，(100) 以至少 $1-\exp\left[-\left(\binom{n}{2}(\frac{1}{2}\sigma_j^2+1)(\sigma_j^2+d_1+1)\right)^{\frac{1}{3}}\right]$ 的概率成立，因此在事件 $\mathcal{B}_n$ 上，(100) 也以至少 $1-\exp\left[-\left(\binom{n}{2}(\frac{1}{2}\sigma_j^2+1)(\sigma_j^2+d_1+1)\right)^{\frac{1}{3}}\right]$ 的概率成立。结合假设 $d/n\to 0$，我们有

$$
\sqrt{\frac{n}{2}}\left|\left\langle(\widehat{\mathcal{P}}_j-\bar{\mathbb{E}}\widehat{\mathcal{P}}_j-\mathcal{L}_j^{[1]})\mathbf{u},\mathbf{v}\right\rangle\right|\cdot\mathbb{I}(\mathcal{B}_n)\xrightarrow{p}0.
$$

E.4 对 (88) 的证明。

在本部分中，我们将使用引理7 来证明

$$
\sqrt{\frac{n}{2}}\left\|\bar{\mathbb{E}}\widehat{\mathcal{P}}_j-\mathcal{P}_j-\mathcal{P}_j\bar{\mathbb{E}}(\mathcal{S}_j^{[1]})\mathcal{P}_j\right\|\xrightarrow{p}0.
$$

根据引理 4,

$$
\begin{aligned}
&\bar{\mathbb{E}}\|\widehat{M}_j^{[1]}-M_j^{[1]}\|+C\|M_j^{[1]}\|\frac{\log n}{n} \\
\leq & C\|M_j^{[1]}\|\left(\sqrt{\frac{r(M_j^{[1]})}{n}}\bigvee\frac{r(M_j^{[1]})}{n}\right)+C\|M_j^{[1]}\|\frac{\log n}{n} \\
\leq & C(\sigma_j^2+1)\left[\frac{\sigma_j^2+d_1+1}{(\frac{1}{2}\sigma_j^2+1)n}+\frac{\log n}{n}\right],
\end{aligned}
$$

在事件 $\mathcal{B}_n$。最后一个不平等成立是因为（98）和（94）。请记住， $\bar{g}_1 \geq \frac{1}{2}\sigma_j^2$（不平等 - 96），由于 $d = o(n)$，对于足够大的 n， $\delta_n(t) \leq \frac{1}{4}\bar{g}_1$。

现在，根据引理7，存在一个常数 $\gamma = \frac{1}{2}$，使得存在 $D := D_{\frac{1}{2}}$在事件$\mathcal{B}_n$上：

$$\begin{aligned}
&\sqrt{\frac{n}{2}}\left\|\bar{\mathbb{E}}\widehat{\mathcal{P}}_j - \mathcal{P}_j - \mathcal{P}_j\bar{\mathbb{E}}(\mathcal{S}_j^{[1]})\mathcal{P}_j\right\| \\
=&D\sqrt{n/2}\frac{\|M_j^{[1]}\|^2}{\bar{g}_1^2}\frac{1}{\sqrt{n/2}}\left(\sqrt{\frac{r(M_j^{[1]})}{n}}\bigvee\sqrt{\frac{\log n}{n}}\right) \\
\leq&4D\left(\frac{\sigma_j^2+1}{\frac{1}{2}\sigma_j^2}\right)^2\left[\sqrt{\frac{\sigma_j^2+d_1+1}{n(\frac{1}{2}\sigma_j^2+1)}}\bigvee\sqrt{\frac{\log n}{n}}\right],
\end{aligned} \tag{101}$$

当 n足够大时。最后一个不等式成立是因为（94）、（96）和（98）。由此即可得到 (88) 该等式。

E.5 对（89）的证明。

注意，对于$\mathbf{u}$， $\mathbf{v} \in \mathbb{H}$,

$$\langle\mathcal{L}_j^{[1]}\mathbf{u},\mathbf{v}\rangle = \frac{2}{n}\sum_{k=1}^{n/2}\left[\langle z_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle z_{jk},\mathcal{C}_j^{[1]}\mathbf{u}\rangle + \langle z_{jk},\mathcal{P}_j\mathbf{u}\rangle\langle z_{jk},\mathcal{C}_j^{[1]}\mathbf{v}\rangle\right], \tag{102}$$

并记住 $\mathbf{h}L_j\mathbf{u},\mathbf{v}\mathbf{i} = \frac{2}{n}\sum_{k=1}^{n/2}[\mathbf{h}y_{jk},\mathbf{P}_j\mathbf{v}\mathbf{i}\mathbf{h}y_{jk},\mathbf{C}_j\mathbf{u}\mathbf{i} + \mathbf{h}y_{jk},\mathbf{P}_j\mathbf{u}\mathbf{i}\mathbf{h}y_{jk},\mathbf{C}_j\mathbf{v}\mathbf{i}],$ ，因此为了证明（89），我们只需证明：对于 $\forall\mathbf{u},\mathbf{v} \in \mathbb{H}$， $\mathbf{k}\mathbf{u}\mathbf{k},\mathbf{k}\mathbf{v}\mathbf{k} \leq 1$,

$$\sqrt{\frac{2}{n}}\sum_{k=1}^{n/2}\left[\langle z_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle z_{jk},\mathcal{C}_j^{[1]}\mathbf{u}\rangle - \langle y_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle y_{jk},\mathcal{C}_j\mathbf{u}\rangle\right] \xrightarrow{p} 0. \tag{103}$$

我们进一步将（103）的证明分为三个步骤。

步骤 1。

$$\begin{aligned}
&\langle z_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle z_{jk},\mathcal{C}_j^{[1]}\mathbf{u}\rangle - \langle y_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle z_{jk},\mathcal{C}_j^{[1]}\mathbf{u}\rangle \\
=&\langle z_{jk}-y_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle z_{jk},\mathcal{C}_j^{[1]}\mathbf{u}\rangle
\end{aligned} \tag{104}$$

但在条件于 $\{\mathscr{X}_{n/2+1}, \mathscr{X}_{n/2+2}, \ldots, \mathscr{X}_n\}$时，$\langle z_{jk}-y_{jk}、\mathcal{P}_j\mathbf{v}\rangle$ 以及$\langle z_{jk}、\mathcal{C}_j^{[1]}\mathbf{u}\rangle$、 $k = 1、\ldots、n^-$均为均值为 0 的高斯随机变量，且不相关，因此相互独立。所以 $\mathbb{E}(\langle z_{jk} -$

y_{jk}、$\mathcal{P}_j\mathbf{v}\rangle\langle z_{jk}$，$\mathcal{C}_j^{[1]}\mathbf{u}\rangle) = 0$。此外，直接计算可得

$$
\begin{aligned}
&\bar{\mathbb{E}}\langle z_{jk} - y_{jk}, \mathcal{P}_j\mathbf{v}\rangle^2 \\
=&\langle \mathbf{u}_j^{(1)}, \mathbf{v}\rangle^2 \bar{\mathbb{E}}\left[\mathscr{X}_k \times_1 \mathbf{u}_j^{(1)} \times_2 \widehat{\mathbf{u}}_j^{(2)[2]} \cdots \times_p \widehat{\mathbf{u}}_j^{(p)[2]} - \mathscr{X}_k \times_1 \mathbf{u}_j^{(1)} \times_2 \mathbf{u}_j^{(2)} \cdots \times_p \mathbf{u}_j^{(p)}\right]^2 \\
=&\langle \mathbf{u}_j^{(1)}, \mathbf{v}\rangle^2 \sigma_j^2 \prod_{q=2}^{r}\left(1 - \langle \widehat{\mathbf{u}}_j^{(q)[2]}, \mathbf{u}_j^{(q)}\rangle\right)^2 + \langle \mathbf{u}_j^{(1)}, \mathbf{v}\rangle^2 \left\|\widehat{\mathbf{u}}_j^{(2)[2]} \otimes \cdots \otimes \widehat{\mathbf{u}}_j^{(p)[2]} - \mathbf{u}_j^{(2)} \otimes \cdots \otimes \mathbf{u}_j^{(p)}\right\|^2 \\
=&\langle \mathbf{u}_j^{(1)}, \mathbf{v}\rangle^2 \sigma_j^2 \prod_{q=2}^{r}\left(1 - \langle \widehat{\mathbf{u}}_j^{(q)[2]}, \mathbf{u}_j^{(q)}\rangle\right)^2 + \langle \mathbf{u}_j^{(1)}, \mathbf{v}\rangle^2 \left(2 - 2\prod_{q=2}^{r}\langle \widehat{\mathbf{u}}_j^{(q)[2]}, \mathbf{u}_j^{(q)}\rangle\right) \\
\leq&\sigma_j^2\left(\frac{1}{4}a_n^4\right)^{p-1} + 2 - 2\left(1 - \frac{1}{2}a_n^2\right)^{p-1}
\end{aligned} \tag{105}
$$

因为 $\|\widehat{\mathbf{u}}_j^{(q)[2]} - \mathbf{u}_j^{(q)}\| \leq a_n$ 这意味着 $1 - \langle \widehat{\mathbf{u}}_j^{(q)[2]}, \mathbf{u}_j^{(q)}\rangle \leq \frac{1}{2}a_{n^\circ -}^2$此外，我们还有

$$
\bar{\mathbb{E}}\left[\langle z_{jk}, \mathcal{C}_j^{[1]}\mathbf{u}\rangle\right]^2 = \langle \mathcal{C}_j^{[1]} M_j^{[1]} \mathcal{C}_j^{[1]}\mathbf{u}, \mathbf{u}\rangle, \tag{106}
$$

其中

$$
\mathcal{C}_j^{[1]} M_j^{[1]} \mathcal{C}_j^{[1]} := \sum_{l\neq j} \frac{\tilde{\sigma}_l^2}{\left(\tilde{\sigma}_j^2 - \tilde{\sigma}_l^2\right)^2}\mathbf{u}_l^{(1)} \otimes \mathbf{u}_l^{(1)} + \frac{1}{\tilde{\sigma}_j^4}\mathcal{P}_{-}. \tag{107}
$$

如前所证，$(1 - a_n^2)^{\frac{p}{2}}\sigma_j^2 \leq \tilde{\sigma}_j^2$，且 $\max_{l\neq j}\{\tilde{\sigma}_l^2\} \leq Aa_n \to 0$，并且当 n 足够大时，$\tilde{\sigma}_j^2 - \tilde{\sigma}_l^2 \geq \frac{1}{2}\sigma_j^2$，因此当 n 足够大时，我们有

$$
\left\|\mathcal{C}_j^{[1]} M_j^{[1]} \mathcal{C}_j^{[1]}\right\| \leq \frac{1}{\left[(1 - a_n^2)^{\frac{p}{2}}\sigma_j^2\right]^4}. \tag{108}
$$

因此，

$$
\begin{aligned}
&\bar{\mathbb{E}}\left[\langle z_{jk} - y_{jk}, \mathcal{P}_j\mathbf{v}\rangle\langle z_{jk}, \mathcal{C}_j^{[1]}\mathbf{u}\rangle\right]^2 \\
\leq&\frac{1}{\left[(1 - a_n^2)^{\frac{p}{2}}\sigma_j^2\right]^4}\left[\sigma_j^2\left(\frac{1}{4}a_n^4\right)^{p-1} + 2 - 2\left(1 - \frac{1}{2}a_n^2\right)^{p-1}\right].
\end{aligned} \tag{109}
$$

步骤 2。

$$
\begin{aligned}
&\langle y_{jk}, \mathcal{P}_j\mathbf{v}\rangle\langle z_{jk}, \mathcal{C}_j^{[1]}\mathbf{u}\rangle - \langle y_{jk}, \mathcal{P}_j\mathbf{v}\rangle\langle y_{jk}, \mathcal{C}_j^{[1]}\mathbf{u}\rangle \\
=&\langle y_{jk}, \mathcal{P}_j\mathbf{v}\rangle\langle z_{jk} - y_{jk}, \mathcal{C}_j^{[1]}\mathbf{u}\rangle
\end{aligned} \tag{110}
$$

但条件于 $\{\mathscr{X}_{n/2+1}, \mathscr{X}_{n/2+2}, \ldots, \mathscr{X}_n\}$, $\mathrm{h}y_{jk}$, $\mathrm{P}\mathcal{P}j\mathrm{vi}$ 和 $\mathrm{h}z_{jk} - y_{jk}$, $\mathrm{C}_j^{[1]}\mathrm{ui}$，$k = 1, \ldots, n^{-}$，它们是均值为 0 的高斯随机变量，且不相关，因此相互独立。所以 $\mathbb{E}(\mathrm{h}z_{jk} - y_{jk}, \mathrm{P}\mathcal{P}j\mathrm{vih}z_{jk}, \mathrm{C}_j^{[1]}\mathrm{ui}) = 0$。

此外,

$$
\begin{aligned}
&\bar{\mathbb{E}}\langle z_{jk}-y_{jk},\mathcal{C}_j^{[1]}\mathbf{u}\rangle^2\\
=&\sum_{l\neq j}\sigma_l^2\langle\mathcal{C}_j^{[1]}\mathbf{u},\mathbf{u}_l^{(1)}\rangle^2\prod_{q=2}^{r}\left(\langle\widehat{\mathbf{u}}_j^{(q)[2]},\mathbf{u}_l^{(q)}\rangle\right)^2\\
&\quad+\left\|\mathcal{C}_j^{[1]}\mathbf{u}\right\|^2\cdot\left\|\widehat{\mathbf{u}}_j^{(2)[2]}\otimes\cdots\otimes\widehat{\mathbf{u}}_j^{(p)[2]}-\mathbf{u}_j^{(2)}\otimes\cdots\otimes\mathbf{u}_j^{(p)}\right\|^2\\
\leq&\left\|\mathcal{C}_j^{[1]}\right\|^2\max_{l\neq j}\left|\sigma_l^2\langle\mathbf{u}_l^{(2)},\widehat{\mathbf{u}}_j^{(2)[2]}\rangle\right|\sqrt{1-\langle\mathbf{u}_j^{(2)},\widehat{\mathbf{u}}_j^{(2)[2]}\rangle^2}+\left\|\mathcal{C}_j^{[1]}\right\|^2\left(2-2\left(1-\frac{1}{2}a_n^2\right)^{p-1}\right)
\end{aligned}
\tag{111}
$$

由于

$$
\mathcal{C}_j^{[1]}=\sum_{l\neq j}\frac{1}{\tilde{\sigma}_j^2-\tilde{\sigma}_l^2}\mathbf{u}_l^{(1)}\otimes\mathbf{u}_l^{(1)}+\frac{1}{\tilde{\sigma}_j^2}\mathcal{P}_-,
\tag{112}
$$

正如已经证明的，当 n, $\tilde{\sigma}_j^2-\tilde{\sigma}_l^2\geq\frac{1}{2}\sigma_j^2$ 足够大时,- 因此当 n 足够大时，我们有

$$
\left\|\mathcal{C}_j^{[1]}\right\|\leq\frac{2}{\sigma_j^2}.
\tag{113}
$$

因为 $\|\widehat{\ }\mathbf{u}_j^{(q)[2]}-\mathbf{u}_j^{(q)}\|\leq a_n$，且 $\max_{l\neq j}\left|\sigma_l^2\langle\mathbf{u}_l^{(2)},\ \widehat{\ }\mathbf{u}_j^{(2)[2]}\rangle\right|\leq A$,

$$
\begin{aligned}
&\bar{\mathbb{E}}\langle z_{jk}-y_{jk},\mathcal{C}_j^{[1]}\mathbf{u}\rangle^2\\
\leq&\frac{2}{\sigma_j^2}\left[Aa_n+2-2\left(1-\frac{1}{2}a_n^2\right)^{p-1}\right]
\end{aligned}
\tag{114}
$$

且 $E\langle y_{jk},\ \mathcal{P}_j\mathbf{v}\rangle^2=(\sigma_j^2+1)\langle\mathbf{u}_j^{(1)},\ v\rangle^2,$，因此

$$
\bar{\mathbb{E}}\left[\langle y_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle z_{jk}-y_{jk},\mathcal{C}_j^{[1]}\mathbf{u}\rangle\right]^2\leq(\sigma_j^2+1)\frac{2}{\sigma_j^2}\left[Aa_n+2-2\left(1-\frac{1}{2}a_n^2\right)^{p-1}\right]
\tag{115}
$$

步骤 3。

$$
\begin{aligned}
&\langle y_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle y_{jk},\mathcal{C}_j^{[1]}\mathbf{u}\rangle-\langle y_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle y_{jk},\mathcal{C}_j\mathbf{u}\rangle\\
=&\langle y_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle y_{jk},(\mathcal{C}_j^{[1]}-\mathcal{C}_j)\mathbf{u}\rangle
\end{aligned}
\tag{116}
$$

但条件于 $\{\mathscr{X}_{n/2+1},\mathscr{X}_{n/2+2},\ldots,\mathscr{X}_n\}$，$\mathrm{h}y_{jk}$、$\mathrm{P}_j\mathrm{v}i$ 和 $\mathrm{h}y_{jk}$、$(\mathrm{C}_j^{[1]}-\mathrm{C}_j)\mathrm{u}i$，$k=1,\ldots,n$ 是均值为 0 的高斯随机变量，且不相关，因此相互独立。所以 $\bar{\mathbb{E}}(\mathrm{h}y_{jk},\mathrm{P}_j\mathrm{v}i\mathrm{h}y_{jk},(\mathrm{C}_j^{[1]}-\mathrm{C}_j)\mathrm{u}i)=0$。

此外,

$$
\mathcal{C}_j^{[1]}-\mathcal{C}_j=\sum_{l\neq j}\left(\frac{1}{\tilde{\sigma}_j^2-\tilde{\sigma}_l^2}-\frac{1}{\sigma_j^2}\right)\mathbf{u}_l^{(1)}\otimes\mathbf{u}_l^{(1)}+\left(\frac{1}{\tilde{\sigma}_j^2}-\frac{1}{\sigma_j^2}\right)\mathcal{P}_--\frac{1}{\sigma_j^2}\left(I_{\mathbb{H}}-I_{d_1}\right)
\tag{117}
$$

$$
\begin{aligned}
&\bar{\mathbb{E}}\langle y_{jk},(\mathcal{C}_j^{[1]}-\mathcal{C}_j)\mathbf{u}\rangle^2\\
=&\left\|\left[\sum_{l\neq j}\left(\frac{1}{\tilde{\sigma}_j^2-\tilde{\sigma}_l^2}-\frac{1}{\sigma_j^2}\right)\mathbf{u}_l^{(1)}\otimes\mathbf{u}_l^{(1)}+\left(\frac{1}{\tilde{\sigma}_j^2}-\frac{1}{\sigma_j^2}\right)\mathcal{P}_-\right]\mathbf{u}\right\|^2
\end{aligned}
\tag{118}
$$

但正如已经证明的，$(1-a_n^2)^{\frac{p}{2}}\sigma_j^2\leq\tilde{\sigma}_j^2$，且 $\max_{l\neq j}\{\tilde{\sigma}_l^2\}\leq Aa_n\to 0$，并且当$n$足够大时，$\tilde{\sigma}_j^2-\tilde{\sigma}_l^2\geq\frac{1}{2}\sigma_j^2$，因此当 n足够大时，我们有

$$
\begin{aligned}
&\left\|\sum_{l\neq j}\left(\frac{1}{\tilde{\sigma}_j^2-\tilde{\sigma}_l^2}-\frac{1}{\sigma_j^2}\right)\mathbf{u}_l^{(1)}\otimes\mathbf{u}_l^{(1)}+\left(\frac{1}{\tilde{\sigma}_j^2}-\frac{1}{\sigma_j^2}\right)\mathcal{P}_-\right\|\\
=&\max_{l\neq j}\left(\frac{1}{\tilde{\sigma}_j^2-\tilde{\sigma}_l^2}-\frac{1}{\sigma_j^2}\right)\bigvee\left(\frac{1}{\tilde{\sigma}_j^2}-\frac{1}{\sigma_j^2}\right)\\
=&\max_{l\neq j}\left(\frac{1}{\tilde{\sigma}_j^2-\tilde{\sigma}_l^2}-\frac{1}{\sigma_j^2}\right)\\
\leq&\frac{1}{(1-a_n^2)^{\frac{p}{2}}\sigma_j^2-Aa_n}-\frac{1}{\sigma_j^2}
\end{aligned}
\tag{119}
$$

并且 $E\langle y_{jk},\ \mathcal{P}_j\mathbf{v}\rangle^2=(\sigma_j^2+1)\langle\mathbf{u}_j^{(1)},\ v\rangle^2$，因此

$$
\bar{\mathbb{E}}\left[\langle y_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle y_{jk},(\mathcal{C}_j^{[1]}-\mathcal{C}_j)\mathbf{u}\rangle\right]^2\leq(\sigma_j^2+1)\left(\frac{1}{(1-a_n^2)^{\frac{p}{2}}\sigma_j^2-Aa_n}-\frac{1}{\sigma_j^2}\right).
\tag{120}
$$

结合（109）、（115）、（120），我们可得，对于事件 $\mathcal{B}_n$，当 n 足够大时，

$$
\begin{aligned}
&\bar{\mathbb{E}}\left(\langle z_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle z_{jk},\mathcal{C}_j^{[1]}\mathbf{u}\rangle-\langle y_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle y_{jk},\mathcal{C}_j\mathbf{u}\rangle\right)^2\\
\leq&\frac{1}{\left[(1-a_n^2)^{\frac{p}{2}}\sigma_j^2\right]^4}\left[\sigma_j^2\left(\frac{1}{4}a_n^4\right)^{p-1}+2-2\left(1-\frac{1}{2}a_n^2\right)^{p-1}\right]\\
&+(\sigma_j^2+1)\frac{2}{\sigma_j^2}\left[Aa_n+2-2\left(1-\frac{1}{2}a_n^2\right)^{p-1}\right]+(\sigma_j^2+1)\left(\frac{1}{(1-a_n^2)^{\frac{p}{2}}\sigma_j^2-Aa_n}-\frac{1}{\sigma_j^2}\right).
\end{aligned}
$$

注意，条件于$\{\mathscr{X}_{n/2+1},\ \mathscr{X}_{n/2+2},\ \ldots,\ \mathscr{X}_n\}$，

$$
\langle z_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle z_{jk},\mathcal{C}_j^{[1]}\mathbf{u}\rangle-\langle y_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle y_{jk},\mathcal{C}_j\mathbf{u}\rangle
$$

彼此独立，因此

$$
\begin{aligned}
&\bar{\mathbb{E}}\left(\sqrt{\frac{2}{n}}\sum_{k=1}^{n/2}\left[\langle z_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle z_{jk},\mathcal{C}_j^{[1]}\mathbf{u}\rangle-\langle y_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle y_{jk},\mathcal{C}_j\mathbf{u}\rangle\right]\right)^2\\
\leq&\frac{1}{\left[(1-a_n^2)^{\frac{p}{2}}\sigma_j^2\right]^4}\left[\sigma_j^2\left(\frac{1}{4}a_n^4\right)^{p-1}+2-2\left(1-\frac{1}{2}a_n^2\right)^{p-1}\right]\\
&+(\sigma_j^2+1)\frac{2}{\sigma_j^2}\left[Aa_n+2-2\left(1-\frac{1}{2}a_n^2\right)^{p-1}\right]+(\sigma_j^2+1)\left(\frac{1}{(1-a_n^2)^{\frac{p}{2}}\sigma_j^2-Aa_n}-\frac{1}{\sigma_j^2}\right),
\end{aligned}
$$

对于事件$\mathcal{B}_n$，当 n 足够大时。结合 $\lim_n a_n=0$，可得

$$
\begin{aligned}
&\mathbb{E}\left\{\left(\sqrt{\frac{2}{n}}\sum_{k=1}^{n/2}\left[\langle z_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle z_{jk},\mathcal{C}_j^{[1]}\mathbf{u}\rangle-\langle y_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle y_{jk},\mathcal{C}_j\mathbf{u}\rangle\right]\right)\cdot\mathbb{I}(\mathcal{B}_n)\right\}^2\\
=&\mathbb{E}\left\{\bar{\mathbb{E}}\left(\sqrt{\frac{2}{n}}\sum_{k=1}^{n/2}\left[\langle z_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle z_{jk},\mathcal{C}_j^{[1]}\mathbf{u}\rangle-\langle y_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle y_{jk},\mathcal{C}_j\mathbf{u}\rangle\right]\right)^2\cdot\mathbb{I}(\mathcal{B}_n)\right\}\\
\to&\ 0.
\end{aligned}
$$

So

$$
\left(\sqrt{\frac{2}{n}}\sum_{k=1}^{n/2}\left[\langle z_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle z_{jk},\mathcal{C}_j^{[1]}\mathbf{u}\rangle-\langle y_{jk},\mathcal{P}_j\mathbf{v}\rangle\langle y_{jk},\mathcal{C}_j\mathbf{u}\rangle\right]\right)\cdot\mathbb{I}(\mathcal{B}_n)\xrightarrow{p}0,
$$

从而证明了（103），并且（89）也随之成立。