arXiv:2304.12551v1 [math.ST] 25 Apr 2023

关于谱嵌入一致性的研究

Ruofei Zhao、Songkai Xue 和 Yuekai Sun

University of Michigan 1085 S University, Ann Arbor, MI 48109

e-mail: rfzhao@umich.edu; sxue@umich.edu; yuekai@umich.edu

摘要：本文研究由某些相似度矩阵的主特征向量得到的谱嵌入向其总体对应项的收敛。我们选择以一致（而非平均）的意义研究这种收敛，并强调这一选择的优势。利用牛顿-坎托罗维奇定理及泛函分析中的其他工具，我们首先为不变子空间的标准正交基建立一个一般性的扰动结果。随后，我们将这一一般结果应用于归一化谱聚类。借助丰富的索伯列夫空间相关文献，并利用希尔伯特空间中的一些集中性结果，我们得以证明归一化谱聚类中谱嵌入一致性误差的有限样本误差界。

2020年数学主题分类： 主要分类 62H30, 47A55。**K**

谱聚类、一致性、牛顿-坎托罗维奇定理、索伯列夫空间、泛函分析、希尔伯特空间中的集中性。

关键词：谱 Ruofei Zhao、Songkai Xue 和 Yuekai Sun

目录

1 引言.21.1 主要结果.31.2 相关文献.51.3 主要贡献.71.4 论文结构.72 预备知识与记号.7

2.1 算子理论.72.2 线性算子的谱理论.92.3 函数空间.92.4 再生核希尔伯特空间（RKHS）.10

3 谱嵌入的一致误差界.10

3.1 第一步：不变子空间的方程刻画 . . 133.2 第二步：使用牛顿-坎托罗维奇定理求解方程 143.3 第三步：证明不变空间是主特征空间 163.4 第四步：界定一致性误差.183.5 第五步：整合所有部分.20

4 归一化谱聚类的应用.214.1 证明概述.234.2 第一部分：Sobolev空间 $\mathcal{H}^s$ .24

4.3 第二部分：算子差界限.254.4 希尔伯特空间上的集中性.264.5 检查一般理论的条件.28 4.6 第五部分：由 $\widehat{V}_1$ .30
5 讨论.33 参考文献.34 A 证明.36 A.1 引理 3.1 的证明.36 A.2 证明定理 3.5 时所用引理的证明.37 A.3 引理4.11 的证明.38 B 核主成分分析的应用.39 B.1 证明概述.42 B.2 第一部分：空间 $\mathcal{L}_{HS}(\mathcal{H})$ 以及 $\mathcal{H}$ 中的随机元素和 $\mathcal{L}_{HS}(\mathcal{H})$ 42 B.3 第二部分：希尔伯特空间 $\mathcal{L}_{HS}(\mathcal{H})$ 中的集中性.43 B.4 第三部分：检验一般理论的条件.43 B.5 第四部分：整合所有要素.44

1. 引言

谱方法是现代统计学中的基础工具。对于聚类或分类等统计学习任务，可以先用谱方法将数据转换为特征，再基于这些特征执行相应任务。在过去二十年里，谱方法已广泛应用于图像分割 [33], 新颖性检测 [18], 社区检测 [13], 生物信息学 [17], ，其有效性部分归功于它能够揭示数据中的潜在低维结构。

谱嵌入之所以得名，是因为嵌入是由正定矩阵的谱分解构造而成的。例如，在归一化谱聚类中，归一化拉普拉斯嵌入 [27]，$\Phi_n: \{x_i\}_{i=1}^n \to \mathbb{R}^K$ 定义为

$$\Phi_n(x_i)^T = e_i^T V, \quad i \in [n], \tag{1.1}$$

其中，$\{x_i\}_{i=1}^n$ 为观测值，$e_i \in \mathbb{R}^n$除第 i个位置为 1 外，其余位置均为 0，K 为嵌入所需的维数，$V \in \mathbb{R}^{n\times K}$ 的列为归一化拉普拉斯矩阵的主特征向量。如前所述，谱嵌入仅定义在训练数据中的点上，但可以通过样本外扩展[4, 41]对训练数据之外的点进行计算。其他谱方法的例子包括 Isomap [36], 拉普拉斯 [3] Hessian特征映射 [14], 和扩散映射 [10]。

由于下游流程将嵌入作为输入，因此嵌入必须具备某些一致性性质，以确保最终输出的质量。具体而言，我们提出

- 在大样本极限下，数据的嵌入表示是否会“收敛”到某些总体层面表示?
- 如果嵌入表示确实收敛，那么它们是以何种意义收敛的?

尽管关于特征值和谱投影收敛性的结果很多，但在一般情形下直接讨论嵌入表示收敛性的结果却很少。唯一的例外是 von Luxburg、Belkin 和 Bousquet [41]。这是现有文献中的一个空白，因为下游应用的输入是嵌入表示，而不是谱投影或特征值。在本文中，我们回答这两个问题并给出直接答案——我们证明，样本层面嵌入在酉变换意义下，一致收敛于其总体对应算子。通过考虑多维嵌入并允许存在非单特征值，我们改进了 von Luxburg、Belkin 和 Bousquet 的结果[41] 。

为具体说明我们的结果在实际中的应用，让我们回到谱聚类。归一化拉普拉斯嵌入的总体对应算子表示为

$$\Psi(x)^T = \begin{bmatrix} f_1(x) & \dots & f_K(x) \end{bmatrix},$$

其中， f_1、...、 f_K 是归一化拉普拉斯算子的主特征函数[41, 30]。正如 von Luxburg、Belkin 和 Bousquet [41], 所示，归一化拉普拉斯矩阵具有一个算子对应形式，我们将其称为经验归一化拉普拉斯算子。设 $f_{n,1}$、...、 $f_{n,K}$ 是该算子的主特征函数，并定义嵌入

$$\Psi_n(x)^T = \begin{bmatrix} \hat{f}_{n,1}(x) & \dots & \hat{f}_{n,K}(x) \end{bmatrix}. \tag{1.2}$$

嵌入 Ψ_n 在样本点上与 Φ_n 重合，即对于所有 $\{x_i\}_{i=1}^n$， $\Psi_n(x_i) = \Phi_n(x_i)$。我们将证明，样本层面的嵌入一致收敛于其总体对应算子:

$$\sup\{d(\Psi_n(x), \Psi(x)) : x \in \mathcal{X}\} \xrightarrow{p} 0, \tag{1.3}$$

其中， d 是定义在 $\mathbb{R}^K$上的某个度量。这意味着， Φ_n 一致收敛于 Ψ 在样本点上的限制。

1.1. 主要结果

本节将以非正式的方式陈述我们的结果。这些结果将在后续各节中得到精确表述和证明。

我们的第一个主要结果涉及扰动对算子不变子空间的影响。它为建立一致性类型的结果提供了一套通用方法。尽管在统计学和机器学习中，我们主要处理实值函数，但我们的主要谱扰动结果是针对复值函数表述的。这一选择在技术上较为方便，因为复数是代数闭的，而实数不是。在

该结果的大多数应用中，复值函数只取实数值。

设 $\mathcal{H}$ 是一个复希尔伯特空间，其元素是在定义域 $\mathcal{X}$上有界的复-~值连续函数。设 T、T是从$\mathcal{H}$ 到 $\mathcal{H}$ 的两个算子，并且它们在希尔伯特–施密特范数下彼此接近。设 $\{f_i\}_{i=1}^K$ 是 T的前 K个特征函数， $\{\tilde{f}_i\}_{i=1}^K$ 是 $\widetilde{T}$的前 K个特征函数。只要对 $\{f_i\}_{i=1}^K$ 和 $\{\tilde{f}_i\}_{i=1}^K$ 进行适当归一化，我们就期望 $\{f_i\}_{i=1}^K$ 在某个酉变换下接近 $\{\tilde{f}_i\}_{i=1}^K$。事实确实如此，我们的第一个结果对此作出了如下刻画。

结果 1（一致性的一般方法）。定义 $V_1:\ \mathbb{C}^K \to \mathcal{H}$ 为$V_1\alpha = \sum_{i=1}^K \alpha_i f_i$，以及 $\widetilde{V}_1:\ \mathbb{C}^K \to \mathcal{H}$ 为 $\widetilde{V}_1\alpha = \sum_{i=1}^K \alpha_i \tilde{f}_i$。存在仅依赖于 T的常数~、C_1、$C_2 > 0$，使得只要 $\|T - T\|_{HS} \le C_1$，我们就有

$$\inf\{\|V_1 - \widetilde{V}_1 Q\|_{2\to\infty} : Q \in \mathbb{U}^K\} \le C_2 \|\widetilde{T} - T\|_{HS}, \tag{1.4}$$

其中，$\mathbb{U}^K$ 是 $\mathbb{C}^{K\times K}$中酉矩阵的空间，而算子 A: $\mathbb{C}^K \to \mathcal{H}$ 的 $2\to\infty$范数定义为

$$\|A\|_{2\to\infty} = \sup\{\|A\alpha\|_\infty : \alpha \in \mathbb{C}^K, \|\alpha\|_2 = 1\}.$$

不难看出，(1.4)与(1.3)之间存在对应关系：V_1 是~的类比；Ψ；V_1 是 Ψ_n 的类比；二到无穷范数保证了收敛的一致性，而距离度量 d 则用于衡量经过酉变换后的欧几里得距离。[1] 这一观察说明，将 $\inf\{\|V_1 - \widetilde{V}_1 Q\|_{2\to\infty}$ 命名为 $Q \in \mathbb{U}^K\}$ 一致性误差是合理的。还值得一提的是，常数 C_2 与 K阶和 $K+1$阶特征值之间的特征值间隙度量 T成反比。这进一步说明，研究整个主特征空间的收敛，而不是像 von Luxburg、Belkin 和 Bousquet 那样研究各个特征子空间的收敛，是有道理的 [41]。前者不仅更具一般性，也更符合实际，而且还能得到更好的常数。在许多应用中，最大的几个特征值通常聚集在一起，但它们与谱中其余部分之间存在较大的间隙。因此，很难分别估计相应的特征函数，但可以在酉变换意义下较容易地将它们作为整体进行估计。

结果 1 提供了一种证明一致性的通用方法：我们只需在适当的范数下界定样本层面算子与总体对应算子之间的差异。结果 1 的证明本身也很有意义。我们通过直接求解算子方程来识别不变子空间，并借助 Newton-Kantorovich 定理刻画其解。这种方法的主要优点在于，它克服了使用非酉不变范数时传统方法的局限性。

[1]同一特征值对应的归一化特征函数仅在酉变换意义下确定。

我们的第二个主要结果是归一化谱聚类中嵌入的有限样本一致误差界。令 $C_b(\mathcal{X})$ 表示 $\mathcal{X}$ 上有界连续复值函数构成的空间。定义 $V_1 : \mathbb{C}^K \to C_b(\mathcal{X})$ 为 $V_1\alpha = \sum_{i=1}^K \alpha_i f_i$，其中 f_1、……、f_K 是归一化拉普拉斯算子的主要实值特征函数。定义 $\widehat{V}_{n,1} : \mathbb{C}^K \to C_b(\mathcal{X})$ 为 $\widehat{V}_{n,1}\alpha = \sum_{i=1}^K \alpha_i \widehat{f_{n,i}}$，其中 f_n、$_i$ 如 (1.2) 中所定义且为实值函数。应用结果 1，我们得到

结果 2（归一化谱聚类的一致性）。在适当条件下，存在常数 $C_4, C_5 > 0$，它们与 n 及样本的随机性无关；只要样本量 $n \geq C_4\tau$ 满足某个 $\tau > 1$，我们就有

$$\inf\{\|V_1 - \widehat{V}_{n,1}Q\|_{2\to\infty} : Q \in \mathbb{U}^K\} \leq C_5 \frac{\sqrt{\tau}}{\sqrt{n}},$$

其概率至少为 $1 - 8e^{-\tau}$。

尽管结果 2 是结果 1 的一个应用，但其证明绝非简单。主要的技术难点在于建立希尔伯特-施密特算子的集中界。结果 2 表明，在适当条件下，收敛速率为 $\mathcal{O}(\frac{1}{\sqrt{n}})$（相差一个对数因子）。此外，在聚类的背景下，一致性的概念能够对聚类结果的正确性提供更有力的保证。例如，在谱聚类中，点是根据其嵌入进行聚类的。一致收敛意味着所有点的嵌入都接近其总体对应算子。只要嵌入的误差足够小，就有可能证明所有点都被正确聚类。如果 $\widehat{}$ 嵌入仅“依均值收敛”，则无法做到这一点：$\|V_1 - V_{n,\ 1}Q\|_{2\to L^2(\mathcal{X},\ \mathbb{P})} \to 0$。

1.2. 相关文献

与我们的结果最为相关的是 von Luxburg、Belkin 和 Bousquet [41] 以及 Rosasco、Belkin 和 Vito [29]的工作。对于归一化谱聚类，von Luxburg、Belkin 和 Bousquet [41] 证明了样本层面算子的特征值和谱投影收敛到相应的总体算子对象。他们还证明了对应特征值的重数为一的特征函数一致收敛到相应的总体特征函数。我们的结果与他们的工作方向相同，因为我们也研究特征函数的一致收敛。我们通过同时考虑多个特征函数并允许存在非单特征值，改进了他们关于一致收敛的结果。在非归一化谱聚类的背景下，Rosasco、Belkin 和 Vito [29] 研究了样本层面算子的有序谱与总体算子的有序谱之间 l^2距离的收敛速率，并推导出与前 K 个特征值相关的样本层面和总体层面谱投影之间偏差的有限样本界。他们还得到了非对称归一化图拉普拉斯的有限样本谱投影误差界。我们的工作与他们相关，因为两者都研究

主特征空间的收敛性，而且我们的许多集中性结果都得益于他们的工作。与此同时，这两项工作也有很大不同。首先，我们所说的收敛是特征函数的一致性，而他们考察的是谱投影之间诱导的再生核希尔伯特空间范数。据我们所知，从一组结果推导出另一组结果并非易事。其次，我们研究的是归一化对称图拉普拉斯，而他们研究的是未归一化图拉普拉斯矩阵和非对称归一化图拉普拉斯。

经验算子/矩阵的谱性质与其总体对应算子谱性质之间的一般关系，也在其他背景下得到研究。在 Koltchinskii 和 Gin´e [22], 的研究中，证明了积分算子及其经验版本的有序谱几乎必然在 l^2-距离下趋于零，当且仅当核是平方可积的。在更强的条件下，他们还得到了收敛速率和分布极限。在 Koltchin- skii [21], 中，作者通过证明由谱投影诱导的二次型满足大数定律和中心极限定理，拓展了他们自己的结果。Mendelson 和 Pajor[25] 以及 Mendelson 和 Pajor [26], 继续研究谱的收敛性，他们将有序谱之间的各种距离度量与 i.i.d. 秩一算子样本均值偏离其总体均值的程度联系起来。核主成分分析（KPCA）文献中也研究了类似问题。例如，在 Shawe-Taylor 等人 [31] 以及 Blanchard、Bousquet 和 Zwald [5], 的研究中，考察了经验核矩阵最大 K 个特征值之和以及除最大 K 个特征值之外的所有特征值之和的集中性，因为这些部分和与 KPCA 的重构误差密切相关。在 Zwald 和 Blanchard [42], 的研究中，推导了一个有限样本误差界，用于衡量经验协方差算子的主特征空间投影算子与总体协方差算子的主特征空间投影算子之间的差异。我们指出，本段提及的结果均未直接讨论嵌入的一致性，也没有研究由度矩阵归一化的核矩阵。

与我们的模型无关结果不同，谱方法的性质也已在特定模型设定下得到研究。例如，Rohe、Chatterjee 和 Yu [28] 以及 Lei 和 Rinaldo [23] 研究了随机块模型下图拉普拉斯算子的谱收敛性质，以及从社区成员恢复角度考察的谱聚类一致性。当数据从非参数分布的有限混合中采样时，Shi、Belkin 和 Yu [32] 研究了总体层面的积分算子的主特征函数和特征向量如何反映聚类信息；Schiebinger、Wainwright 和 Yu [30] 研究了由归一化谱聚类生成的嵌入样本的几何结构，并表明，当混合成分的重叠较小且样本量较大时，不同聚类的嵌入样本近似正交。我们指出，迄今为止本节提及的所有结果中，核函数都是固定的。关于流形上的图拉普拉斯算子与拉普拉斯-贝尔特拉米算子之间的关系，以及在核函数自适应选取时谱聚类的性质，读者可参阅 Trillos 等人 [38, 39, 37] 的系列工作

及其中的参考文献。

最后，近期文献还研究了矩阵特征向量和特征子空间的逐元素或逐行分析。对于一般情形，Fan、Wang 和 Zhong$\ell_{2\to\infty}$ 、Cape、Tang 和 Priebe [16], 以及 Damle 和 Sun[7], 分别给出了界 [12], ，其中前两项针对矩形矩阵，最后一项针对对称矩阵。当对真实矩阵和扰动矩阵施加概率假设时，Cape、Tang 和 Priebe [8], 、Abbe 等人[1], 以及 Mao、Sarkar 和 Chakrabarti [24] 利用随机矩阵的结构 $\ell_{2\to\infty}$ ，针对各种任务获得了更强的界。与这些文献相比，我们指出，本文提供了一个有助于 $\ell_{2\to\infty}$ 线性算子扰动理论的确定性界，并且该界可应用于统计学中的许多问题（例如谱聚类和核主成分分析），从而帮助刻画单个样本的谱嵌入。

1.3. 主要贡献

我们将主要贡献概括为三个方面，并按其出现顺序列出。首先，我们证明，牛顿-坎托罗维奇定理为研究局部扰动对算子不变子空间的影响提供了一种通用方法。这一结果对于从事谱扰动理论研究的学者而言，可能具有独立的研究价值。其次，我们通过一致性误差研究嵌入的收敛性，并提出了一种建立非渐近一致性结果的通用方法。该方法能够同时处理多个特征函数，且不局限于简单特征值。第三，我们将该方法应用于归一化谱聚类，并针对谱嵌入的一致性误差，给出有限样本误差界的一种新证明。

1.4. 论文结构

本文其余部分的组织如下：第 2 节回顾相关的数学预备知识；第 3 节给出结果 1 的精确表述及证明；第 4 节给出结果 2 的精确表述及证明；第 5 节讨论与本文结果相关的各种问题；一些次要引理的证明及另一项应用则放在附录中。

2. 预备知识与符号约定

本节将介绍本文中反复用到的若干基本概念和预备结果。针对特定章节的更具技术性的结果，将在后文需要时再行介绍。

2.1. 算子理论

我们假设读者熟悉巴拿赫空间、希尔伯特空间、线性算子、算子范数和算子谱等基本概念。从

现在起，我们用 $\mathbb{K}$ 表示实数域或复数域中的任一个，用 $\mathcal{Y}_1$、$\mathcal{Y}_2$ 表示同一域 $\mathbb{K}$上的巴拿赫空间，用 $\mathcal{H}_1$、$\mathcal{H}_2$ 表示同一域 $\mathbb{K}$上的希尔伯特空间。

我们首先想强调一下线性算子定义中的一个细微之处。对于线性算子 $A:\mathcal{Y}_1 \to \mathcal{Y}_2$，我们采用 Kato [20]的约定，允许 A 仅定义在线性流形 $\mathcal{Y}_1$中，[2] 记为 $D(A)$。[3]我们称 $D(A)$ 为 A 的定义域，并且可以自然地将 A 的值域定义为$R(A):=\{Ay \mid y \in D(A)\}$。至于 $\mathcal{Y}_1$和$\mathcal{Y}_2$，我们分别称它们为定义域空间和值域空间。

对于线性算子 $A:\mathcal{Y}_1 \to \mathcal{Y}_2$，如果 $\sup_{\|y\|_{\mathcal{Y}_1}=1}\|Ay\|_{\mathcal{Y}_2}<\infty$，我们称 A 为有界的；当 A 有界时，我们定义其算子范数为 $\|A\|:=\sup_{\|y\|_{\mathcal{Y}_1}=1}\|Ay\|_{\mathcal{Y}_2}$。在全文中，当 $\|\cdot\|$ 没有下标时，默认表示算子范数。我们用 $\mathcal{L}(\mathcal{Y}_1,\mathcal{Y}_2)$ 表示从 $\mathcal{Y}_1$ 到$\mathcal{Y}_2$的所有有界线性算子组成的空间。当 $\mathcal{Y}_1=\mathcal{Y}_2$ 时，我们将 $\mathcal{L}(\mathcal{Y}_1,\mathcal{Y}_1)$ 简写为 $\mathcal{L}(\mathcal{Y}_1)$。如果有界集在 $\mathcal{Y}_1$ 中经过 A 映射后的像的闭包是紧的，我们称 A 为紧算子。众所周知，紧算子是有界的。

对于有界线性算子 $A:\mathcal{H}_1 \to \mathcal{H}_2$，定义其伴随算子 $A^*:\mathcal{H}_2 \to \mathcal{H}_1$为从 $\mathcal{H}_2$ 到 $\mathcal{H}_1$ 的唯一算子，并满足 $\langle Af, g\rangle_{\mathcal{H}_2}=\langle f, A^*g\rangle_{\mathcal{H}_1}$，其中$\forall f \in \mathcal{H}_1$、$\forall g \in \mathcal{H}_2$。这里，我们使用 $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ 表示希尔伯特空间 $\mathcal{H}$中的内积。A^* 的一个基本性质是 $\|A\|_{\mathcal{H}_1\to\mathcal{H}_2}=\|A^*\|_{\mathcal{H}_2\to\mathcal{H}_1}$，其中范数为算子范数，并且我们使用 $\mathcal{H}_1 \to \mathcal{H}_2$ 记号来明确指定定义域空间和值域空间。当 $\mathcal{H}_1=\mathcal{H}_2$、$A$ 时，若 A 等于其伴随算子 A^*，则称其为自伴算子；若对于任意 $f \in \mathcal{H}_1$、$\langle Af, f\rangle_{\mathcal{H}_1} \geq 0$成立，则称 A 为正算子。

如果一个希尔伯特空间存在由可数多个元素组成的基，我们就称它是可分的。若有界线性算子 $A:\mathcal{H}_1 \to \mathcal{H}_2$ 满足 $\sum_{i\in\mathcal{I}}\|Ae_i\|^2_{\mathcal{H}_2}<\infty$，其中 $\{e_i:i\in\mathcal{I}\}$ 是 $\mathcal{H}_1$的标准正交基，则称其为希尔伯特-施密特算子。我们用 $HS(\mathcal{H}_1,\mathcal{H}_2)$ 表示从 $\mathcal{H}_1$到 $\mathcal{H}_2$的所有希尔伯特-施密特算子构成的空间；关于内积$\langle A,B\rangle_{HS}:=\sum_{i\in\mathcal{I}}\langle Ae_i, Be_i\rangle_{\mathcal{H}_2}$，该空间也是一个希尔伯特空间。我们用 $\|\cdot\|_{HS}$表示由此内积诱导的范数，并注意到所有希尔伯特-施密特算子都是紧算子。还要指出，希尔伯特–施密特范数强于算子范数，即$\|A\| \leq \|A\|_{HS}$；而且希尔伯特–施密特范数与算子范数相容，具体来说，对于任意希尔伯特-施密特算子 A 和有界算子 B，它们的乘积 AB 和 BA 都是希尔伯特-施密特算子，且其希尔伯特–施密特范数满足

$$\|AB\|_{HS} \leq \|A\|_{HS}\|B\|,$$
$$\|BA\|_{HS} \leq \|B\|\|A\|_{HS}.$$

[2]在 Kato 中，[20], 线性流形只是仿射子空间的同义词。[3]例如，在 Ciarlet 中，[9], 并不作这样的区分。

2.2. 线性算子的谱理论

在本小节中，我们令 $\mathbb{K} = \mathbb{C}$。设 $A: \mathcal{H}_1 \to \mathcal{H}_1$ 是一个有界线性算子。与矩阵的情形类似，若对于某个特征向量$f \in \mathcal{H}_1$，有 $\lambda \in \mathbb{C}$，则称 $\lambda \in \mathbb{C}$ 是 A 的一个特征值，

$$Af = \lambda f \quad \text{and} \quad f \neq 0.$$

换言之，当零空间 $N(\lambda I - A)$ 不是 $\{0\}$ 时，λ 是一个特征值。我们将与 λ 相关联的 $N(\lambda I - A)$ 称为特征子空间，而 $N(\lambda I - A)$ 的维数称为 λ 的几何重数。A 的谱定义为 $\sigma(A) := \mathbb{C} \setminus \rho(A)$，其中 $\rho(A)$ 是预解集

$$\rho(A) := \{\lambda \in \mathbb{C} \mid (\lambda I - A)^{-1} \in \mathcal{L}(\mathcal{H}_1)\}.$$

特征值属于谱，但 $\sigma(A)$ 通常不仅包含特征值。如果 A 是紧算子，$\sigma(A)$ 具有以下结构：$\sigma(A) \setminus \{0\}$ 是一个由孤立特征值组成的可数集合，每个特征值的几何重数均为有限，并且 $\sigma(A)$ 的唯一可能聚点是 0。如果 A 是自伴的，则所有特征值都必须是实数。如果 A 是正算子，则其所有特征值都是实数且非负。因此，对于任意紧正自伴算子，我们可以将 A 的非零特征值排列成一个非递增的正数序列，[4] 并将每个特征值重复其几何重数所对应的次数。

线性算子谱理论中的另一个重要事实涉及谱投影。设 $\Gamma \subset \rho(A)$ 是一条闭的简单可求长曲线。假设 $\sigma(A)$ 中位于 Γ 内部的部分包含有限个特征值 $\lambda_1, \lambda_2, \ldots, \lambda_K$。那么，可以定义投影 $P: \mathcal{H}_1 \to \mathcal{H}_1$，它投影到 $\{\lambda_i\}_{i=1}^K$ 的特征子空间的直和上，即 $\bigoplus_{i=1}^K N(\lambda_i I - A)$。撇开技术细节不谈，该投影可表示为以下围道积分

$$P = \frac{1}{2\pi i} \int_\Gamma (\gamma I - A)^{-1} d\gamma.$$

2.3. 函数空间

设 $\mathcal{X}$ 是 $\mathbb{R}^p$ 的一个有界开子集，我们现在定义几个接下来要用到的函数空间。将有界连续函数空间 $C_b(\mathcal{X})$ 定义为

$$C_b(\mathcal{X}) := \{f \mid f : \mathcal{X} \to \mathbb{C} \text{ is a bounded continous function}\}.$$

可以证明，$\|f\|_\infty := \sup_{x \in \mathcal{X}} |f(x)|$ 是 $C_b(\mathcal{X})$ 上的一个范数，而 $C_b(\mathcal{X})$ 关于该无穷范数构成一个巴拿赫空间。

我们还可以定义复值平方可积函数的空间$L^2(\mathcal{X}, \mu)$。设（$\mathcal{X}$，$\mathcal{B}$，μ）是一个测度空间，其中 $\mathcal{B}$是勒贝格 σ-代数，μ是一个测度，则 $L^2(\mathcal{X}, \mu)$定义为可测

[4]因为最大特征值受 A 的算子范数所界定。

函数，使得

$$\int_{\mathcal{X}} |f(x)|^2 d\mu < \infty.$$

事实上， $L^2(\mathcal{X},\mu)$ 关于内积是一个希尔伯特空间

$$\langle f, g\rangle_{L^2(\mathcal{X},\mu)} := \int_{\mathcal{X}} f\bar{g} d\mu.$$

我们还定义了 l^2，即由复数组成的平方可和无穷序列空间。众所周知， l^2 关于内积是一个复希尔伯特空间

$$\langle u, v\rangle_{l^2} := \sum_{i=1}^{\infty} u_i \bar{v_i}.$$

2.4. 再生核希尔伯特空间（RKHS）

设 $\mathcal{X}$ 是 $\mathbb{R}^p$ 的子集， $\mathcal{H}$ 是一组函数 $f:\ \mathcal{X} \to \mathbb{C}$。假设 $\mathcal{H}$ 关于某个内积 $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ 是一个希尔伯特空间。如果在 $\mathcal{H}$ 中，所有点求值泛函都是有界的，即

$$|f(x)| \le C_x \|f\|_{\mathcal{H}} \quad \forall f \in \mathcal{H},$$

其中 C_x 是依赖于 x 的某个常数，那么可以证明存在唯一的共轭对称正定核函数 $k:\ \mathcal{X}\times\mathcal{X} \to \mathbb{C}$，使得满足以下再生性质：

$$f(x) = \langle f, k(\cdot, x)\rangle.$$

核 k 被称为再生核，而 $\mathcal{H}$ 被称为再生核希尔伯特空间（RKHS）。

如果对于任意 $n \in \mathbb{N}^+$、任意 x_1、x_2、……、$x_n \in \mathcal{X}$以及任意 ξ_1、ξ_2、……、$\xi_n \in \mathbb{C}$，二次型 k为

$$\sum_{i,j=1}^{n} k(x_i, x_j)\bar{\xi_i}\xi_j \tag{2.1}$$

非负的，则称核函数为正定的。任何再生核希尔伯特空间的核函数都是正定的。

3. 谱嵌入的一致误差界

在本节中，我们证明上一节所述的第一个结果。首先给出假设和记号。设 $\mathcal{X}$ 是 $\mathbb{R}^p$ 的子集， $\mathbb{P}$ 是一个概率测度，其密度函数的支撑集包含于 $\mathcal{X}$。令 $L^2(\mathcal{X},\mathbb{P})$表示 $\mathcal{X}$ 上的复值平方可积函数空间，并令 $\mathcal{H}$ 为 $L^2(\mathcal{X},\mathbb{P})$的一个子空间。我们假设 $\mathcal{H}$配备了自身的内积$\langle\cdot,\cdot\rangle_{\mathcal{H}}$，且相对于该内积构成一个希尔伯特空间。我们还要求

$\mathcal{H}$满足：对于每个 $h \in \mathcal{H}$（它是 $L^2(\mathcal{X},\mathbb{P})$中的一个等价类），该类中都存在一个代表函数 h'，使得 $h' \in C_b(\mathcal{X})$。由于 $\mathrm{supp}(\mathbb{P}) = \mathcal{X}$，$h'$ 是唯一的，因此我们可以在 $\mathcal{H}$ 上定义无穷范数，令 $\|h\|_\infty := \|h'\|_\infty$。我们要求在 $\mathcal{H}$上，由 $\mathcal{H}$内积诱导的范数（记作 $\|\cdot\|_{\mathcal{H}}$）强于无穷范数；也就是说，存在一个常数 $C_{\mathcal{H}} > 0$，使得对所有 $f \in \mathcal{H}$ 都有 $\|f\|_\infty \le C_{\mathcal{H}}\|f\|_{\mathcal{H}}$。[5]

设T和$\widetilde{T}$是从$\mathcal{H}$到$\mathcal{H}$的两个希尔伯特-施密特算子，T表示二者之差。假设 T 的所有特征值（按几何重数计）可以排列成一个非递增（可能无限）的非负实数序列$\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_K > \lambda_{K+1} \ge \ldots \ge 0$，且 λ_K与λ_{K+1}之间存在正间隙。假设 T 的特征值也可以排列成一个非递增的非负实数序列。不过，我们不假设 T存在任何特征值间隙。

设 $\{f_i\}_{i=1}^K \subset \mathcal{H}$ 是与特征值 λ_1、.....、λ_K对应的特征函数。我们假设 $\{f_i\}_{i=1}^K$ 的选取方式使它们构成 $L^2(\mathcal{X},\mathbb{P})$中的一组正交向量。接着，我们选取 $\{f_i\}_{i=K+1}^\infty$，使得 $\{f_i\}_{i=1}^\infty$构成 $L^2(\mathcal{X},\mathbb{P})$的一组完备正交基。定义 $V_1\colon \mathbb{C}^K \to L^2(\mathcal{X},\mathbb{P})$，使其满足 $V_1\alpha = \sum_{i=1}^K \alpha_i f_i$；并定义 $V_2\colon l^2 \to L^2(\mathcal{X},\mathbb{P})$，使其满足 $V_2\beta = \sum_{i=1}^\infty \beta_i f_{K+i}$。定义它们的伴随算子$V_1^*$、$V_2^*$，其所依据的内积是 $\mathbb{C}^K$、l^2 以及 $L^2(\mathcal{X},\mathbb{P})$上的标准内积。由于$\{f_i\}_{i=1}^K \subset \mathcal{H}$，我们也可以将 $\mathcal{H}$ 视为 V_1 (V_1^*)的值域（定义域）空间。V_1 的确切值域空间将由上下文说明。

当T必然具有特征值间隙，在这种情况下，$\widetilde{T}$ 的主 K 维不变子空间是良定义的。我们选取 $\{\tilde{f}_i\}_{i=1}^K$ 作为 $L^2(\mathcal{X},\mathbb{P})$ 中的一组正交归一向量，使其张成 $\widetilde{T}$ 的主不变子空间，并定义$\widetilde{V}_1\colon \mathbb{C}^K \to L^2(\mathcal{X},\mathbb{P})$ 为 $\widetilde{V}_1\alpha = \sum_{i=1}^K \alpha_i \tilde{f}_i$。

最后但同样重要的是，定义 $V_2^{-1}\mathcal{H} := \{l \in l^2 : V_2 l \in \mathcal{H}\}$ 以及 $V_2^*\mathcal{H} := \{V_2^* h : h \in \mathcal{H}\}$。直观地说，它们是在 V_2 的基下，$\mathcal{H}$ 中函数的“坐标空间”。使用这些坐标可以简化我们的记号。关于 $V_2^{-1}\mathcal{H}$ 和 $V_2^*\mathcal{H}$ 的以下事实成立（证明见附录）。

引理 3.1。假设 f_1，.....，$f_K \in \mathcal{H}$，则有

1. 集合 $V_2^{-1}\mathcal{H}$ 等于集合 $V_2^*\mathcal{H}$；2. $V_2^{-1}\mathcal{H}$ 是 l^2 的子空间；相对于由 $\mathcal{H}$ 诱导的内积，它也是一个希尔伯特空间

$$(b_1, b_2)_{V_2^{-1}\mathcal{H}} := \langle V_2 b_1, V_2 b_2\rangle_{\mathcal{H}};$$

3. $V_1 \in \mathcal{L}(\mathbb{R}^K,\mathcal{H})$、$V_1^* \in \mathcal{L}(\mathcal{H},\mathbb{R}^K)$、$V_2 \in \mathcal{L}(V_2^{-1}\mathcal{H},\mathcal{H})$和 $V_2^* \in \mathcal{L}(\mathcal{H},V_2^{-1}\mathcal{H})$,

[5]这实际上意味着 $\mathcal{H}$ 是一个再生核希尔伯特空间。但由于证明中完全没有用到再生性质，我们认为将 $\mathcal{H}$ 描述为再生核希尔伯特空间没有必要。

其算子范数满足

$$\|V_1\|_{2\to\mathcal{H}} \le \sqrt{K}\max_{i\in[K]}\|f_i\|_{\mathcal{H}}, \quad \|V_1^*\|_{\mathcal{H}\to 2} \le C_{\mathcal{H}}\sqrt{K},$$
$$\|V_2\|_{V_2^{-1}\mathcal{H}\to\mathcal{H}} = 1, \quad \|V_2^*\|_{\mathcal{H}\to V_2^{-1}\mathcal{H}} \le 1 + C_{\mathcal{H}}K\max_{i\in[K]}\|f_i\|_{\mathcal{H}}.$$

由于引理的第 1 项，我们无需区分$V_2^{-1}\mathcal{H}$和 $V_2^*\mathcal{H}$；将二者都记为 $\tilde{l}^2$。为使记号简洁，对任意 i、$j\in\{1,2\}$，定义$\sim$ $\sim$ $\sim$ $\sim$ $T_{ij} = V_i^*TV_j$；例如， T_{21}是 $V_2^*TV_1$的简写。

我们还需要以下量来定义结果 1 中的常数。令Γ为矩形的边界

$$\Big\{\lambda\in\mathbb{C}\mid \frac{\lambda_K+\lambda_{K+1}}{2} \le re(\lambda) \le \|T\|_{\mathcal{H}\to\mathcal{H}}+1, |im(\lambda)|\le 1\Big\}. \tag{3.1}$$

令 $l(\Gamma)$表示 Γ 的长度，并定义

$$\eta = \frac{1}{\sup_{\lambda\in\Gamma}\|(\lambda I - A)^{-1}\|_{op}}, \tag{3.2}$$

该值必然有限。定义一个衡量谱分离度的量

$$\delta := \mathrm{sep}(T_{11},T_{22}) := \inf\Big\{\big\|T_{22}Y - YT_{11}\big\|_{HS} \;\Big|\; Y\in\mathcal{L}(\mathbb{C}^K,\tilde{l}^2), \|Y\|_{HS}=1\Big\}.$$

可以合理地预期，特征值间隙越大，$\mathrm{sep}(T_{11},T_{22})$也越大。当 T仅有 K 个特征值，或从 $\mathcal{H}$ 到 $\mathcal{H}$是自伴时，可以证明分离度 $\mathrm{sep}(T_{11},T_{22})$以下界于特征值间隙 $\lambda_K-\lambda_{K+1}$。定义常数

$$C_3 := \max\Big\{C_{\mathcal{H}}, 1+C_{\mathcal{H}}\|V_1\|_{2\to\mathcal{H}}, \|V_1\|_{2\to\mathcal{H}}(1+C_{\mathcal{H}}\|V_1\|_{2\to\mathcal{H}})\Big\}.$$

现在我们可以陈述本节的主要定理了。

定理 3.2（一致性的通用方案）。在上述假设下，只要 $E := T - T$ 作为从 $\mathcal{H}$ 到 $\mathcal{H}$ 的算子，其 Hilbert-Schmidt 范数为 $\|E\|_{HS}\le C_1$，一致性误差就满足

$$\inf\{\|V_1 - \widetilde{V}_1Q\|_{2\to\infty} : Q\in\mathbb{U}^K\} \le C_2\|E\|_{HS}.$$

这里，C_1、$C_2>0$ 是两个与 T的选择无关的常数，定义如下：

$$C_1 := \frac{1}{C_3}\min\Big\{\frac{\lambda_K-\lambda_{K+1}}{8}, \frac{1}{2}, \frac{\delta}{4}, \frac{\delta}{4C_{\mathcal{H}}}, \frac{C_3\eta^2}{\eta + l(\Gamma)/2\pi}\Big\}, \tag{3.3}$$
$$C_2 := \frac{4C_3C_{\mathcal{H}}(\|V_1\|_{2\to\infty}+1)}{\delta}. \tag{3.4}$$

我们指出，由于 C_2 与 δ成反比，因此研究主特征空间整体的收敛性是有益的。若分别考察各个特征子空间，由于主导特征值可能聚集在一起，且 δ较小，因此每个特征子空间的收敛速度都较慢；但若将它们作为整体考察，则 δ 会更大，从而收敛更快，因为主导特征值与谱的其余部分之间有明显间隔。

本节其余部分旨在证明定理3.2。证明策略是~将 V_1Q 表示为算子方程的解，并直接~界定 $\|V_1 - V_1Q\|_{2\to\infty}$。我们分五个步骤给出证明。第一步，我们~将 T 的不变子空间表示为二次算子方程的解。第二步，我们应用牛顿-坎托罗维奇定理，证明当扰动较小时，该方程确实存在解。第三步，我们引入一些附加条件，以保证第二步得到的不变子空间是主不变子空间。第四步，我们直接界定~误差项 $\|V_1 - V_1Q\|_{2\to\infty}$。第五步，我们汇总所有结果并证明定理3.2。Stewart [35] 也采用了类似的方法来研究矩阵的不变子空间。

3.1. 第一步：不变子空间的方程刻画

在本节中，我们的目标是找到一个 $Y \in \mathcal{L}(\mathbb{C}^K, \tilde{l}^2)$，使得 $V_1 + V_2Y \in \mathcal{L}(\mathbb{C}^K, \mathcal{H})$ 的值域是 $\widetilde{T}$ 的不变子空间。事实证明，任何满足以下二次算子方程的 Y 都足够。

命题 3.3. 只要 $Y \in \mathcal{L}(\mathbb{C}^K, \tilde{l}^2)$ 满足方程

$$\widetilde{T}_{21} + \widetilde{T}_{22}Y = Y\widetilde{T}_{11} + Y\widetilde{T}_{12}Y, \tag{3.5}$$

$V_1 + V_2Y \in \mathcal{L}(\mathbb{C}^K, \mathcal{H})$ 的值域就是 $\widetilde{T}$ 的不变子空间。

证明。首先，我们注意到（3.5）是 $\mathcal{L}$（$\mathbb{C}^K$，$\tilde{l}^2$）中的一个定义良好的算子方程。~这可由我们的假设 $T \in \mathcal{L}$（$\mathcal{H}$）以及引理 3.1 的第 3 项看出。接下来，我们断言方程（3.5）蕴含[6]

$$\widetilde{T}(V_1 + V_2Y) = (V_1 + V_2Y)(\widetilde{T}_{11} + \widetilde{T}_{12}Y), \tag{3.6}$$

~这表明 $V_1 + V_2Y$的值域在 T作用下是不变的。为证明这一断言，注意

$$\begin{aligned}\widetilde{T}(V_1 + V_2Y) &= (V_1V_1^* + V_2V_2^*)\widetilde{T}(V_1 + V_2Y)\\ &= V_1\widetilde{T}_{11} + V_1\widetilde{T}_{12}Y + V_2\widetilde{T}_{21} + V_2\widetilde{T}_{22}Y\\ &= V_1\widetilde{T}_{11} + V_1\widetilde{T}_{12}Y + V_2Y\widetilde{T}_{11} + V_2Y\widetilde{T}_{12}Y\\ &= (V_1 + V_2Y)(\widetilde{T}_{11} + \widetilde{T}_{12}Y).\end{aligned}$$

□

[6]我们不区分 $L^2(\mathcal{X}, \mathbb{P})$ 中的等式与 $\mathcal{H}$ 中的等式，因为二者是等价的。

3.2. 第二步：使用牛顿-坎托罗维奇定理求解方程

~将 T的不变子空间刻画为方程 (3.5) 的解之后，我们应用牛顿-坎托罗维奇定理证明方程 (3.5) 存在解。当函数本身及其一阶和二阶导数满足某些条件时，牛顿-坎托罗维奇定理可以构造一个从巴拿赫空间到巴拿赫空间的函数的根。该构造是算法式的：这个根是通过 Newton-Raphson 方法求根所生成的一系列迭代的极限点。我们使用的牛顿-坎托罗维奇定理的确切版本出自 Karow 和 Kressner 的附录 [19]。

定理 3.4（牛顿-坎托罗维奇）。设 $\mathcal{E}$、$\mathcal{Z}$ 是巴拿赫空间，且 $F:\mathcal{Z}\to\mathcal{E}$ 在 $Z\in\mathcal{Z}$ 的一个足够大的邻域 Ω 内二次连续可微。假设存在一个线性算子 $\mathbb{T}:\mathcal{Z}\to\mathcal{E}$，它有连续逆算子 $\mathbb{T}^{-1}$，并满足以下条件：

$$\|\mathbb{T}^{-1}(F(Z))\|_{\mathcal{Z}}\le a, \tag{3.7}$$

$$\|\mathbb{T}^{-1}\circ F'(Z)-I\|_{op}\le b, \tag{3.8}$$

$$\|\mathbb{T}^{-1}\circ F''(\widetilde{Z})\|_{op}\le c,\qquad \forall\widetilde{Z}\in\Omega. \tag{3.9}$$

若 $b<1$ 且 $h:=\frac{ac}{(1-b)^2}<\frac{1}{2}$，则存在 $F(Z_E)=0$ 的一个解 Z_E，使得

$$\|Z_E-Z\|_{\mathcal{Z}}\le r_0\qquad with\qquad r_0:=\frac{2a}{(1-b)(1+\sqrt{1-2h})}.$$

现在我们可以证明下面的命题了。该命题指出，当$\|E_{11}\|_{HS}$、$\|E_{21}\|_{HS}$、$\|E_{22}\|_{HS}$、$\|E_{12}\|_{HS}$ 相对于 sep(T_{11}、T_{22}) 足够小时，方程（3.5）有解。

命题 3.5。设 $\delta:=sep(T_{11}、T_{22})$ 且 $s_E:=\delta-\|E_{22}\|_{HS}-\|E_{11}\|_{HS}$。当 $s_E>0$ 且 $\frac{\|E_{21}\|_{HS}\|E_{12}\|_{HS}}{s_E^2}<\frac{1}{4}$时，存在 $Y\in\mathcal{L}(\mathbb{C}^K、\tilde{l}^2)$，满足 $\|Y\|_{HS}\le\frac{2\|E_{21}\|_{HS}}{s_E}$，使方程 (3.5) 成立。

命题的证明 *3.5*。整理后，(3.5) 等价于

$$E_{21}+(T_{22}+E_{22})Y-Y(T_{11}+E_{11})-YE_{12}Y=0. \tag{3.10}$$

令 $\mathcal{E}:=\mathcal{L}(\mathbb{C}^K,\tilde{l}^2)$ 表示从 $\mathbb{C}^K$到 $\tilde{l}^2$的有界线性算子空间。由于 $\mathbb{C}^K$是有限维的，从 $\mathbb{C}^K$到 $\tilde{l}^2$的任意线性算子都是有界的，并且是 Hilbert-Schmidt 算子。因此，我们可以将 Hilbert-Schmidt范数作为 $\mathcal{E}$ 上的默认范数，而 $\mathcal{E}$ 关于此范数是一个希尔伯特空间。这一事实还使我们能够定义 $f:\mathcal{E}\mapsto\mathcal{E}$为 $f(Y):=E_{21}+(T_{22}+E_{22})Y-Y(T_{11}+E_{11})-YE_{12}Y$，以及$\mathcal{T}:\mathcal{E}\mapsto\mathcal{E}$为 $\mathcal{T}(Y):=T_{22}Y-YT_{11}$。注意到 $\mathcal{H}$ 在 T下的像 E仍然位于 $\mathcal{H}$中，我们可以验证 $\mathcal{T}$和 f 确实定义良好。

我们断言 $\delta>0$ 且 $\mathcal{T}$是单射并且是满射，并将其证明留给一个引理。这意味着 $\mathcal{T}$是可逆的，其中$\|\mathcal{T}^{-1}\|_{op}=\frac{1}{\mathrm{sep}(T_{11},T_{22})}\le\frac{1}{\delta}$。

现在我们准备验证牛顿-坎托罗维奇定理的三个假设。

(A1):

$$\left\|\mathcal{T}^{-1}(f(0))\right\|_{HS} = \left\|\mathcal{T}^{-1}(E_{21})\right\|_{HS} \leq \left\|\mathcal{T}^{-1}\right\|_{op}\left\|E_{21}\right\|_{HS} \leq \frac{\left\|E_{21}\right\|_{HS}}{\delta} =: a. \tag{3.11}$$

(A2): f 在 Y_0 处的 Fr´echet 导数为

$$\begin{aligned} f'(Y_0)\colon \mathcal{E} &\longrightarrow \mathcal{E} \\ \Delta Y &\longmapsto (T_{22}+E_{22})\Delta Y - \Delta Y(T_{11}+E_{11}) \\ &\quad - \Delta Y E_{12}Y_0 - Y_0 E_{12}\Delta Y. \end{aligned}$$

特别地，当 $Y_0 = 0$ 时，

$$\begin{aligned} f'(0)\colon \mathcal{E} &\longrightarrow \mathcal{E} \\ \Delta Y &\longmapsto (T_{22}+E_{22})\Delta Y - \Delta Y(T_{11}+E_{11}). \end{aligned}$$

因此，

$$\begin{aligned} \left(\mathcal{T}^{-1}\circ f'(0) - I\right)\colon \mathcal{E} &\longrightarrow \mathcal{E} \\ \Delta Y &\longmapsto \mathcal{T}^{-1}(E_{22}\Delta Y - \Delta Y E_{11}). \end{aligned}$$

于是有

$$\begin{aligned} \left\|\mathcal{T}^{-1}\circ f'(0) - I\right\| &= \sup_{\|\Delta Y\|_{HS}=1} \left\|\mathcal{T}^{-1}(E_{22}\Delta Y - \Delta Y E_{11})\right\|_{HS} \\ &\leq \sup_{\|\Delta Y\|_{HS}=1} \left\|\mathcal{T}^{-1}\right\|_{op}\left\|E_{22}\Delta Y - \Delta Y E_{11}\right\|_{HS} \\ &\leq \frac{1}{\delta}\sup_{\|\Delta Y\|_{HS}=1}\left\{\left\|E_{22}\right\|_{HS}\left\|\Delta Y\right\|_{HS} + \left\|\Delta Y\right\|_{HS}\left\|E_{11}\right\|_{HS}\right\} \\ &\leq \frac{\|E_{22}\|_{HS} + \|E_{11}\|_{HS}}{\delta} := b. \end{aligned}$$

(A3): Y_0 处的二阶 Fr´echet 导数是$\mathcal{L}(\mathcal{E},\mathcal{L}(\mathcal{E},\mathcal{E}))$中的线性算子：

$$\begin{aligned} f''(Y_0)\colon \mathcal{E} &\longrightarrow \mathcal{L}(\mathcal{E},\mathcal{E}) \\ \Delta_1 Y &\longmapsto \mathcal{T}_{\Delta_1 Y}, \end{aligned}$$

其中 $\mathcal{T}_{\Delta_1 Y}$ 是

$$\begin{aligned} \mathcal{T}_{\Delta_1 Y}\colon \mathcal{E} &\longrightarrow \mathcal{E} \\ \Delta_2 Y &\longmapsto -\Delta_1 Y E_{12}\Delta_2 Y - \Delta_2 Y E_{12}\Delta_1 Y. \end{aligned}$$

因此，对于每个 $Y_0 \in \mathcal{E}$，二阶导数都是常数，于是我们有：

$$\begin{aligned}\left\|\mathcal{T}^{-1} \circ f''(Y_0)\right\|_{op} &= \sup_{\|\Delta_1 Y\|_{HS}=1} \left\|\mathcal{T}^{-1} \circ \mathcal{T}_{\Delta_1 Y}\right\|_{op} \\ &= \sup_{\|\Delta_1 Y\|_{HS}=1} \sup_{\|\Delta_2 Y\|_{HS}=1} \left\|(\mathcal{T}^{-1} \circ \mathcal{T}_{\Delta_1 Y})(\Delta_2 Y)\right\|_{\mathcal{E}} \\ &\leq \frac{1}{\delta} \sup_{\|\Delta_1 Y\|_{HS}=1} \sup_{\|\Delta_2 Y\|_{HS}=1} \left\|\Delta_1 Y E_{12} \Delta_2 Y + \Delta_2 Y E_{12} \Delta_1 Y\right\|_{\mathcal{E}} \\ &\leq \frac{2\|E_{12}\|_{HS}}{\delta} := c\end{aligned}$$

（结论：）在所有假设均已满足的情况下，我们应用牛顿-坎托罗维奇定理，并得出如下结论。当

$$\begin{aligned}&s_E := \delta - \|E_{22}\|_{HS} - \|E_{11}\|_{HS} > 0 \text{ and} \\ &\frac{\|E_{21}\|_{HS}\|E_{12}\|_{HS}}{s_E^2} < \frac{1}{4},\end{aligned}$$

方程（3.5）有解 Y_E 使得

$$\left\|Y_E\right\| \leq \frac{2\|E_{21}\|_{HS}}{s_E + \sqrt{s_E^2 - 4\|E_{21}\|_{HS}\|E_{12}\|_{HS}}} \leq \frac{2\|E_{21}\|_{HS}}{s_E}. \tag{3.12}$$

□

3.3. 第三步：证明不变子空间是主特征空间

在第二步中，我们得到了一个不变子空间，但无法保证所得到的不变子空间就是 K 维的不变子$\tilde{}$空间 T的主特征空间。在本小节中，我们给出确保这一点的充分条件。当$\|E\|_{HS}$较小时，我们证明以下几件事必然成立：首先，$V_1+V_2Y_E\tilde{}$是 K 维的；其次，T 在该子空间上的限制的特征值包含于区间 $[\lambda_K - \epsilon, \lambda_1 + \epsilon]$，其中 ϵ较小；第三，$\sim T$ 在区间 K 内恰好有 $\sim[\lambda_K - \epsilon, \lambda_{max}$（$T$）$+\epsilon]$个特征值（按几何重数计）。综合这些事实可知，命题3.5 中的不变子空间必然是主特征 K维不变子空间。

第一点不难证明。假设 $V_1 + V_2Y_E$ 的值域维数小于 K，那么存在满足 $\|s\|_2 = 1$ 的 $s \in \mathbb{C}^K$，使得 $(V_1 + V_2Y_E)s = 0$。但是，由于 $\|(V_1 + V_2Y_E)s\|_{L^2} \geq \|V_1 s\|_{L^2} - \|V_2 Y_E s\|_{L^2} \geq 1 - C_{\mathcal{H}}\|V_2 Y_E s\|_{\mathcal{H}} \geq 1 - C_{\mathcal{H}}\|Y_E\|_{HS}$，当 $\|Y_E\|_{HS}$ 很小时，向量 $(V_1 + V_2Y_E)s$ 根本不可能为零。形式化地说，我们有引理 3.6。当 $C_{\mathcal{H}}\|Y_E\|_{HS} < 1$ 时，$V_1 + V_2Y_E$ 的值域是 K 维的。

至于第二点，即确定限制 $\tilde{}$ $\tilde{}$ $\tilde{}$ $\tilde{}$ 的特征值，注意到（3.5）意味着 T 在基 $T_{11}+T_{12}Y_E$ 下的矩阵表示为 $V_1+V_2Y_E$。由于特征值不受基的选择影响，$\tilde{}$ $\tilde{}$我们知道 T 在不变子空间上的特征值就是 $T_{11}+T_{12}Y_E$ 的特征值。接下来，我们回顾一个关于特征值的扰动结果。

引理 3.7。假设 $\sigma(T_{11}+E_{11}+E_{12}Y_E)\subset\mathbb{R}$，则有（加法是集合加法）

$$\sigma(T_{11}+E_{11}+E_{12}Y_E)\subset\sigma(T_{11})+\big[-\|E_{11}+E_{12}Y_E\|,\|E_{11}+E_{12}Y_E\|\big]. \tag{3.13}$$

证明。首先注意到 $T_{11}=\mathrm{diag}((\lambda_1,\lambda_2,\ldots,\lambda_K)^T)$。设 λ 是 $T_{11}+E_{11}+E_{12}Y_E\in\mathbb{C}^{K\times K}$ 的一个实特征值，则存在 $v\in\mathbb{C}^K$，使得 $\|v\|_2=1$，从而 $(T_{11}+E_{11}+E_{12}Y_E)v=\lambda v$。因此可得

$$\inf_{i\in[K]}|\lambda_i-\lambda|\le\Big(\sum_{i=1}^{K}(\lambda_i-\lambda)^2|v_i|^2\Big)^{1/2}=\|(\lambda I-T_{11})v\|=\|(E_{11}+E_{12}Y_E)v\|\le\|E_{11}+E_{12}Y_E\|,$$

这表明 λ 至少位于 $\|E_{11}+E_{12}Y_E\|$ 中的一个 $\{\lambda_i\}_{i=1}^K$内。这等价于断言（3.13）。□

由该引理可知（假设 $\|Y_E\|_{HS}<1$）

$$\sigma(T_{11}+E_{11}+E_{12}Y_E)\subset\big[\lambda_K-\|E_{11}\|_{HS}-\|E_{12}\|_{HS},\lambda_1+\|E_{11}\|_{HS}+\|E_{12}\|_{HS}\big]. \tag{3.14}$$

对于第三点，我们需要 Rosasco、Belkin 和 Vito 的以下结果 [29]（他们的定理20），他们将该结果归功于 Anselone [2] 的开创性工作。

定理 3.8. 设 $A\in\mathcal{L}(\mathcal{H})$ 为紧算子。给定 Λ 中的一组有限个非零特征值，其中 A 为紧算子，设 Γ 为任意一条简单可求长闭曲线（取正方向），其内部包含 Λ，且 $\sigma(A)$ 在 $\setminus\Lambda$ 外部。设 P 为与 Λ 相关的谱投影，即

$$P=\frac{1}{2\pi i}\int_\Gamma(\lambda I-A)^{-1}d\lambda, \tag{3.15}$$

并定义

$$\eta^{-1}=\sup_{\lambda\in\Gamma}\|(\lambda I-A)^{-1}\|_{op}. \tag{3.16}$$

设 B 为另一个满足下列条件的紧算子

$$\|B-A\|_{op}\le\frac{\eta^2}{\eta+l(\Gamma)/2\pi}<\eta, \tag{3.17}$$

其中 $l(\Gamma)$ 为 Γ 的长度，则以下事实成立。

$\hat{}$1. 曲线 Γ 是 B 的预解集的子集，并包围着 B的一组有限个非零特征值 Λ;

2. P的值域维数等于 $\widehat{P}$的值域维数，其中 $\widehat{P} = \frac{1}{2\pi i}\int_{\Gamma}(\lambda I - B)^{-1}d\lambda$。

根据上述定理，我们可以取 Γ 如（3.1）所示，即取为矩形的边界

$$\{\lambda \in \mathbb{C} \mid \frac{\lambda_K + \lambda_{K+1}}{2} \leq re(\lambda) \leq \|T\|_{\mathcal{H}\to\mathcal{H}} + 1, |im(\lambda)| \leq 1\}. \tag{3.18}$$

当 $\|E\|_{op} \leq \|E\|_{HS}$ 足够小时， Γ 恰好包含 $\sim T$ 的前 K 个特征值。综合以上三点，我们得到了 $v_1 + V_2 Y_E$ 的值域 $\widetilde{}$ 成为 T 的主不变子空间的充分条件。

命题 3.9。假设

$$\|Y_E\|_{HS} < \min\left\{1, \frac{1}{C_{\mathcal{H}}}\right\} \tag{3.19}$$

$$\|E_{11}\|_{HS} + \|E_{12}\|_{HS} < \min\left\{\frac{\lambda_K - \lambda_{K+1}}{2}, 1\right\}, \tag{3.20}$$

$$\|E\|_{HS} \leq \min\left\{\frac{\eta^2}{\eta + l(\Gamma)/2\pi}, 1\right\}, \tag{3.21}$$

其中 Γ， η 分别根据（3.18）和（3.16）定义，$\sim V_1 + V_2 Y_E$ 的值域是 T的主不变子空间。

证明。首先，我们对 Γ 的选择恰好包含 K 个$\sim T$的最大特征值。接下来，由于我们假设 T 也只有实特征值，因此引理3.7 适用。因此，由（3.14）、（3.19）和（3.20），我们知道

$$\sigma(T_{11} + E_{11} + E_{12}Y_E) \subset \big(\frac{\lambda_K + \lambda_{K+1}}{2}, \lambda_1 + 1\big),$$

它包含在 Γ中。我们还从（3.19）看到，引理3.6 适用。最后，$\widetilde{}$关于 E 的条件（3.8 适用，因此 T 在 Γ中只有 K 个特征值。由此可知，由 $V_1 + V_2 Y_E$ 导出的不变子空间是 K- $\widetilde{}$维的 T主不变子空间。 □

3.4. 第四步：界定一致性误差

在这一步中，我们界定一致性误差

$$\inf\{\|V_1 - \widetilde{V}_1 Q\|_{2\to\infty} : Q \in \mathbb{U}^K\},$$

$\widetilde{}$其中 V_1 的正交归一列张成了 $\sim$ $\sim$ $\sim T$的主不变子空间。由于我们要求 $V_1^* V_1 = I$，因此需要将 $V_1 + V_2 Y_E$的“列”正交归一化。定义 Y_E^* 为 Y_E 关于 $\mathbb{C}^K$ 的伴随算子，并

l^2。我们可以验证，对于某个 $Q \in \mathbb{U}^K$， $\widetilde{V}_1 Q = (V_1 + V_2 Y_E)(I + Y_E^* Y_E)^{-1/2}$，因为[7]

$$(V_1 + V_2 Y_E)^*(V_1 + V_2 Y_E) = V_1^* V_1 + Y_E^* V_2^* V_2 Y_E = I + Y_E^* Y_E.$$

同时注意，根据假设， $\|\cdot\|_{\mathcal{H}}$ 比 $\|\cdot\|_\infty$更强，即对于所有 $f \in \mathcal{H}$都有 $C_{\mathcal{H}}\|f\|_{\mathcal{H}} \geq \|f\|_\infty$。这意味着对于 $\forall l \in \tilde{l}^2$

$$C_{\mathcal{H}}\|l\|_{\tilde{l}^2} = C_{\mathcal{H}}\|V_2 l\|_{\mathcal{H}} \geq \|V_2 l\|_\infty \geq \|V_2 l\|_{L^2} = \|l\|_{l^2}.$$

其结果是

$$C_{\mathcal{H}}\|Y_E\|_{2\to\tilde{l}^2} \geq \|Y_E\|_{2\to l^2}. \tag{3.22}$$

我们还需要下面这个简便的结果。

引理 3.10。设 $Y\colon\ \mathbb{C}^K \to l_2$ 的算子范数为$\|Y\|_{2\to l^2} < 1$。则

$$\left\|(I + Y^*Y)^{-\frac{1}{2}}\right\|_2 \leq 1, \quad \left\|I - (I + Y^*Y)^{-\frac{1}{2}}\right\|_2 \leq \|Y\|_{2\to l^2}. \tag{3.23}$$

证明。假设$\|Y\|_{2\to l^2} = r < 1$。注意到 $\|Y^*Y\| = r^2 < 1$，因此厄米特矩阵 $I + Y^*Y$ 可逆，且其谱位于 $[1, 1+r^2]$中。因此， $\sigma\big((I+Y^*Y)^{-\frac{1}{2}}\big) \subset [\frac{1}{\sqrt{1+r^2}}, 1]$，所以$\|($ $I + Y^*Y^{-\frac{1}{2}})\| \leq 1$。

同样地，我们有 $\sigma(I - (I + Y^*Y)^{-\frac{1}{2}}) \subset [0, 1 - \frac{1}{\sqrt{1+r^2}}]$。还需验证 $1 - \frac{1}{\sqrt{1+r^2}} \leq r$，这很容易。 □

现在我们有

~命题 **3 11** $\|Y_E\|_{2\to\tilde{l}^2} < 1/C_{\mathcal{H}}$ $V_1 Q = V_1 + V_2 Y_E$ $I +$ 。假设对于某个 $Q \in \mathbb{U}^K$，有 $Y_E^* Y_E$ $^{-1/2}$ ()()。我们有

$$\inf\{\|V_1 - \widetilde{V}_1 Q\|_{2\to\infty} : Q \in \mathbb{O}^K\} \leq C_{\mathcal{H}}(\|V_1\|_{2\to\infty} + 1)\|Y_E\|_{2\to\tilde{l}^2}.$$

证明。 条件 $\|Y_E\|_{2\to\tilde{l}^2} < 1/C_{\mathcal{H}}$ 确保引理 3.10 适用。随后我们直接计算

$$\begin{aligned}
&\inf\{\|V_1 - \widetilde{V}_1 Q\|_{2\to\infty} : Q \in \mathbb{O}^K\} \\
&\leq \left\|V_1 - (V_1 + V_2 Y_E)(I + Y_E^* Y_E)^{-\frac{1}{2}}\right\|_{2\to\infty} \\
&\leq \left\|V_1(I - (I + Y_E^* Y_E)^{-\frac{1}{2}})\right\|_{2\to\infty} + \left\|V_2 Y_E (I + Y_E^* Y_E)^{-\frac{1}{2}}\right\|_{2\to\infty} \\
&\leq \|V_1\|_{2\to\infty}\|Y_E\|_{2\to l^2} + \|V_2 Y_E\|_{2\to\infty} \\
&\leq \|V_1\|_{2\to\infty}\|Y_E\|_{2\to l^2} + C_{\mathcal{H}}\|Y_E\|_{2\to\tilde{l}^2} \\
&\leq C_{\mathcal{H}}(\|V_1\|_{2\to\infty} + 1)\|Y_E\|_{2\to\tilde{l}^2}.
\end{aligned}$$

□

[7]正如我们将从引理 3.10中看到的，$(I + Y_E^* Y_E)^{-1/2}$ 在 $\|Y_E\|_{HS}$ 较小时是良定义的。

3.5. 第五步：整合所有部分

我们将前面的步骤结合起来，证明定理 3.2。为此，我们需要以下引理，该引理将 $\|E_{ij}\|_{HS}$、 i和 $j \in \{1,2\}$ 与 $\|E\|_{HS}$联系起来。

引理 3.12. 设

$$C_3 = \max\Big\{C_{\mathcal{H}}, 1 + C_{\mathcal{H}}\|V_1\|_{2\to\mathcal{H}}, \|V_1\|_{2\to\mathcal{H}}(1 + C_{\mathcal{H}}\|V_1\|_{2\to\mathcal{H}})\Big\},$$

那么对于任意 $i,j \in \{1,2\}$

$$\|E_{ij}\|_{HS} \le C_3\|E\|_{HS}.$$

证明。注意到

$$\|V_i^* E V_j\|_{HS} \le \|V_i^*\|_{op}\|E\|_{HS}\|V_j\|_{op}.$$

代入引理 3.1中第 3 项给出的界，即可得到所述结果。 □

定理 *3.2*的证明。定义 C_1 为

$$C_1 := \frac{1}{C_3}\min\Big\{\frac{\lambda_K - \lambda_{K+1}}{8}, \frac{1}{2}, \frac{\delta}{4}, \frac{\delta}{4C_{\mathcal{H}}}, \frac{C_3\eta^2}{\eta + l(\Gamma)/2\pi}\Big\}.$$

根据假设，我们有 $\|E\|_{HS} \le C_1$ 。根据引理 3.12，该假设意味着

$$\|E_{ij}\|_{HS} \le C_3\|E\|_{HS} < \frac{\delta}{4}$$

对所有 i、 $j \in \{1,2\}$ 均成立。因此

$$s_E = \delta - \|E_{11}\|_{HS} - \|E_{22}\|_{HS} \ge \frac{\delta}{2},$$
$$\frac{\|E_{21}\|_{HS}\|E_{12}\|_{HS}}{s_E^2} < \frac{1}{4}.$$

命题3.5 保证（3.5）有一个解 Y_E ，并且

$$\|Y_E\|_{HS} \le \frac{4C_3}{\delta}\|E\|_{HS} < \min\{\frac{1}{C_{\mathcal{H}}}, 1\}. \tag{3.24}$$

我们检验所选取的 C_1 满足条件（3.21）和（3.20），因此命题 3.9 意味着命题 3.5 中的不变子空间就是主不变子空间。最后，（3.24）意味着命题3.11 的条件得到满足，因此我们有

$$\inf\{\|V_1 - \widetilde{V}_1 Q\|_{2\to\infty} : Q \in \mathbb{U}^K\} \le C_2\|E\|_{HS}$$

其中 $C_2 = 4C_3C_{\mathcal{H}}(\|V_1\|_{2\to\infty} + 1)/\delta$。

□

4. 归一化谱聚类的应用

在谱聚类中，我们从一个子集 $\mathcal{X} \subset \mathbb{R}^p$、一个概率测度 $\mathbb{P}$定义在 $\mathcal{X}$上开始，[8] 以及一个连续对称正定实值核函数k： $\mathcal{X} \times \mathcal{X} \to \mathbb{R}$。在观测到样本 X_1、……、X_n 独立同分布$\sim \mathbb{P}$后，我们构造矩阵$K_n \in \mathbb{R}^{n\times n}$来表示它们的两两相似度：$K_n = \left[\frac{1}{n}k(X_i, X_j)\right]_{i,\,j=1}^{n}$，然后将其归一化以得到归一化拉普拉斯矩阵

$$L_n = D_n^{-\frac{1}{2}} K_n D_n^{-\frac{1}{2}},$$

其中 $d_n = K_n 1_n$ 且 $D_n = \mathrm{diag}(d_n)$ 是度矩阵。[9] 可以证明， L_n 是对称半正定的，因此可以对其进行特征值分解。我们将 L_n 的特征对记为 (λ_k, v_k)，并按特征值降序排列：

$$\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_n \geq 0.$$

在本文中，我们对 L_n 的特征向量进行归一化，使得 $n^{-\frac{1}{2}}\|v_k\|_2 = 1$。谱嵌入矩阵为 $V \in \mathbb{R}^{n\times K}$，其列向量为 v_1、……、v_K。

暂且假设核 k 与 0 保持正距离，并且有上界。 L_n 的算子对应形式如下所示，可以证明它是 $\mathcal{L}(C_b(\mathcal{X}))$ 中的有界线性算子

$$(\widehat{T}_n f)(x) = \int_{\mathcal{X}} \frac{k(x,y)}{d_n(x)^{1/2} d_n(y)^{1/2}} f(y) dP_n(y), \tag{4.1}$$

其中 $d_n(x) = \int_{\mathcal{X}} k(x,y) dP_n(y)$ 是样本度函数。尽管 $\widehat{T}_n$ 被引入为 $\mathcal{L}(C_b(\mathcal{X}))$ 中的算子，我们要指出，对于 T_n 而言，决定性元素 $\widehat{\ }$ 是积分形式，定义域空间和值域空间不必严格限制在 $\widehat{\ }(C_b(\mathcal{X}))$ 中。事实上，我们实际要处理的 T_n 是希尔伯特空间之间的算子；这里选择 $\mathcal{L}(C_b(\mathcal{X}))$ 只是为了便于理解。同样的说明也适用于我们随后定义的其他算子。

算子 $T_n = L_n \circ \rho_n$，其中 L_n 的算子对应形式 $\widehat{T}_n$ 是 $\widehat{\ }(C_b(\mathcal{X}))$ 中，因为 ρ_n 是定义如下的限制算子

$$\rho_n f = \begin{bmatrix} f(X_1) & \ldots & f(X_n) \end{bmatrix}^T.$$

换言之，如果通过$\widehat{\ }$将函数 $f \in C_b(\mathcal{X})$与向量 $v \in \mathbb{C}^n$ 对应起来，那么限制算子 ρ_n 就会 T_n “表现为” L_n。 T_n 和 L_n 的特征值与特征向量（函数）也有如下密切关系。

引理 4.1。 设实值核函数 $k(x, y)$ 是连续的，并且$\widehat{\ }$有上下界： $0 < \kappa_l < k(x, y) < \kappa_u < \infty$。令 T_n 按照（4.1）定义，其中定义域空间和值域空间均为 $C_b(\mathcal{X})$。若

[8] 假设 $\mathbb{P}$ 的底层 σ-代数是勒贝格 σ-代数。[9] 归一化拉普拉斯矩阵通常定义为 $I_n - L_n$，但L_n 与 $I_n - L_n$ 的特征向量相同，因此研究 L_n更为方便。

（λ, f）是 T_n 的非平凡特征对（即 $\lambda \neq 0$），则（λ, $\rho_n f$）是 L_n的特征对。反之，若（λ, v）是 L_n的特征对，则（λ, f），其中

$$\widehat{f}(x) = \frac{1}{\widehat{\lambda} n} \sum_{i=1}^{n} \frac{k(x, X_i)}{d_n(x)^{1/2} d_n(X_i)^{1/2}} v_i, \tag{4.2}$$

是 T_n 的特征对，其中 $f \in C_b(\mathcal{X})$。此外，对 f 的选择满足 $\|f\|_{L^2(\mathcal{X}, \mathbb{P}_n)} = 1$，并且 f对样本点的限制与 v一致，即 $\rho_n f = v$。

证明。设（λ, f）是 T_n的特征对：$T_n f = \lambda f$。我们验证（λ, $\rho_n f$）是 L_n的特征对：

$$L_n \rho_n f = \rho_n T_n f = \rho_n \widehat{\lambda} f = \widehat{\lambda} \rho_n f.$$

反之，若（λ, v）是 L_n 的特征对，我们验证（λ, f）是 T_n的特征对：

$$\begin{aligned}
(\widehat{T}_n \widehat{f})(x) &= \frac{1}{n} \sum_{i=1}^{n} \frac{k(x, X_i)}{d_n(x)^{1/2} d_n(X_i)^{1/2}} \left\{ \frac{1}{\widehat{\lambda} n} \sum_{j=1}^{n} \frac{k(X_i, X_j)}{d_n(X_i)^{1/2} d_n(X_j)^{1/2}} v_j \right\} \\
&= \frac{1}{n} \sum_{i=1}^{n} \frac{k(x, X_i)}{d_n(x)^{1/2} d_n(X_i)^{1/2}} \left\{ \frac{1}{\widehat{\lambda}} [L_n v]_i \right\} \\
&= \frac{1}{n} \sum_{j=1}^{n} \frac{k(x, X_j)}{d_n(x)^{1/2} d_n(X_j)^{1/2}} v_i \\
&= (\widehat{\lambda} \widehat{f})(x).
\end{aligned}$$

还需验证 f 确实属于 $C_b(\mathcal{X})$。为此，注意到由于核函数 k 连续且有上界，我们知道 $k(x, X_j) \in C_b(\mathcal{X})$。由于 k 有下界，我们知道 $d_n(x)$连续，且 $d_n(x) > \kappa_l$，因此 $k(x, X_j)/(d_n(x)^{1/2} d_n(X_i)^{1/2}) \in C_b(\mathcal{X})$。因此，这类项的平均值 $\widehat{f}$ 也属于 $C_b(\mathcal{X})$。 □

T_n 的总体版本是归一化拉普拉斯算子

$$T : C_b(\mathcal{X}) \to C_b(\mathcal{X}) \text{ defined as } (Tf)(x) = \int_{\mathcal{X}} \frac{k(x, y)}{d(x)^{1/2} d(y)^{1/2}} f(y) dP(y),$$

其中 $d(x) = \int_{\mathcal{X}} k(x, y) dP(y)$ 是（总体）度函数。在适当的假设下，可以证明我们可以选择 $\{f_i\}_{i=1}^{K}$，即 T 的前 K 个特征函数，使其为实值函数，并且在 $L^2(\mathcal{X}, P)$ 中正交归一。因此，我们可以定义 $V_1 : \mathbb{C}^K \to C_b(X)$ 为 $V_1 \alpha = \sum_{i=1}^{K} \alpha_i f_i$。类似地，根据（4.2），我们可以用 $\{\widehat{f}_k\}_{k=1}^{K}$ 定义 $\widehat{V}_1$，即 L_n 的前 K 个特征向量的延拓。本节的目标是应用一般理论证明以下结果。

定理 4.2。在下文定义的一般假设下，存在由 $\mathcal{X}$、$\mathbb{P}$、k 确定的 C_4、C_5，使得当样本量 $n \geq C_4\tau$ 满足某个 $\tau > 1$ 时，我们以 $1 - 8e^{-\tau}$ 的置信度有

$$\inf\{\|V_1 - \widehat{V}_1 Q\|_{2\to\infty} : Q \in \mathbb{U}^K\} \leq C_5 \frac{\sqrt{\tau}}{\sqrt{n}}.$$

定理4.2 中提及的一般假设

一般假设。集合 $\mathcal{X}$ 是 $\mathbb{R}^p$ 中具有良好边界的有界连通开集。[10]。概率测度 $\mathbb{P}$ 关于勒贝格测度定义，并且具有密度函数 $p(x)$。此外，存在常数 $0 < p_l < p_u < \infty$，使得 $p_l < p(x) < p_u$ 几乎处处（相对于勒贝格测度）。核 $k(\cdot,\cdot) \in C_b^{p+2}(\mathcal{X}\times\mathcal{X})$ 是对称的、正的，并且存在常数 $0 < \kappa_l < \kappa_u < \infty$，使得对于 $\forall x, y \in \mathcal{X}$，$\kappa_l < k(x,y) < \kappa_u$。将其视为从 $L^2(\mathcal{X},\mathbb{P})$ 到 $L^2(\mathcal{X},\mathbb{P})$ 的算子时，T 的特征值满足 $\lambda_1 \geq \ldots \geq \lambda_K > \lambda_{K+1} \geq \ldots \geq 0$。$T$，$\{f_i\}_{i=1}^K \subset C_b^{p+2}$ $\mathcal{X}$ [11] ( ) 的前 K 个特征函数。

4.1. 证明概述

应用一般理论时，最具挑战性的部分是确定要使用的^正确的希尔伯特空间 $\mathcal{H}$，并证明 $T - T_n$作为从 $\mathcal{H}$ 到 $\mathcal{H}$的算子，其希尔伯特–施密特范数会随着 n 趋于无穷而趋于零。在一般假设下，我们可以将 $\mathcal{H}$ 设为一个次数足够高的Sobolev空间^。至于如何对 $\|T - T_n\|_{HS}$进行界定，我们首先将~T, T_n 分解为三个算子的乘积。定义

$$D^{-1/2} : \mathcal{H} \to \mathcal{H} \text{ as } (D^{-1/2}f)(x) = f(x)/\sqrt{d(x)}, \tag{4.3}$$

$$K : \mathcal{H} \to \mathcal{H} \text{ as } (Kf)(x) = \int_{\mathcal{X}} k(x,y)f(y)dP(y). \tag{4.4}$$

于是 $T = D^{-1/2}KD^{-1/2}$。类似地，我们有 $\widehat{T}_n = D_n^{-1/2}K_nD_n^{-1/2}$，其中$D_n^{-1/2}$，$K_n$ 是 $D^{-1/2}$ 和 K 的样本层面版本，分别使用 d_n 和$\mathbb{P}_n$定义。我们将证明 K_n 集中到 K，且 $D_n^{-1/2}$集中到 $D^{-1/2}$，并^利用三角不等式来界定 $\|T - T_n\|_{HS}$。

尽管一般理论已经完成了大部分工作，但为了完成定理 4.2 的完整证明，我们还必须采取一个额外步骤。在我们的一般~理论中，V_1 的列在 $L^2(\mathcal{X},\ \mathbb{P})$ 中正交归一，并且张成了~的主不变子空间 T。然而，在定理 4.2 中，同一个主不变子空间由以下各列张成

[10]我们需要边界满足准解析条件 [6]，以适用于不等式（4.10）以及 C^∞ 引理 A。2 [15, 11]。我们还需要 $\mathcal{X}$ 满足锥条件 [6]由于这些条件的精确定义非常技术化，且与本文的主要论述无关，我们省略其定义。[11]函数空间 C_b^{p+2} $(\mathcal{X})$ 将在第 4.2节中定义。

: $\widehat{V}_1$的列，而它们仅在 $L^2(\mathcal{X}, \mathbb{P}_n)$ 中正交归一。从直观上说，当 n 很大时，V_1 和 V_1 在经过某个酉变换后大致相同，因此从 V_1 切换到 V_1 不应使一致性误差增加任何数量级。确切的误差界将通过某种一致大数定律得到。

严格的论述将分为五个部分。在第一部分，我们介绍所使用的 Sobolev空间，并阐述其基本性质。在第二部分，我们界定诸如 $\|D_n^{-1/2} - D^{-1/2}\|_{\mathcal{H}\to\mathcal{H}}\widehat{\ }$ 和 $\|K - K_n\|_{HS}$ 之类的算子差的范数，并用它们表示 $\|T - T_n\|_{HS}$。在第三部分，我们调用希尔伯特空间中的集中性结果，并将算子差的范数与样本量联系起来。在第四部分，我们检查一般理论所要求的其余条件，并将前面的所有部分结合起来。在第五部分，我们处理由 V_1 与 V_1 之差所引起的误差，并完成证明。

4.2. 第一部分：Sobolev空间 H^s

首先回顾一下，根据假设，$\mathcal{X}$ 是 $\mathbb{R}^p$的一个具有良好边界的有界连通开子集。给定 $s \in \mathbb{N}$，阶数为 s的 Sobolev空间 $\mathcal{H}^s = \mathcal{H}^s(\mathcal{X})$ 定义为

$$\mathcal{H}^s := \{f \in L^2(\mathcal{X}, dx) \mid D^\alpha f \in L^2(\mathcal{X}, dx), \forall |\alpha| = s\},$$

其中 $D^\alpha f$ 是 f 关于多重指标 α 的（弱）导数，$L^2(\mathcal{X}, dx)$ 是由勒贝格测度下平方可积的复值函数构成的复希尔伯特空间。空间 $\mathcal{H}^s$ 关于内积是一个可分Hilbert空间

$$\langle f, g\rangle_{\mathcal{H}^s} = \langle f, g\rangle_{L^2(\mathcal{X},dx)} + \sum_{|\alpha|=s} \langle D^\alpha f, D^\alpha g\rangle_{L^2(\mathcal{X},dx)}.$$

令 $C_b^s(\mathcal{X})$ 表示复值连续有界函数的集合，其中所有不超过 s 阶的导数均存在且为连续有界函数。空间 $C_b^s(\mathcal{X})$ 关于范数是一个巴拿赫空间

$$\|f\|_{C_b^s} = \|f\|_\infty + \sum_{|\alpha|=s} \|D^\alpha f\|_\infty.$$

由于 $\mathcal{X}$有界，我们知道 $C_b^s(\mathcal{X}) \subset \mathcal{H}^s$ 且 $\|f\|_{\mathcal{H}^s} \le C_s \|f\|_{C_b^s}$其中 C_s是一个仅依赖于 s 的常数。我们还根据 Sobolev嵌入定理（参见 Burenkov [6]第 4.6 章）可知，对于 l，$m \in \mathbb{N}$ 当 $l - m > p/2$ 时，我们有

$$\mathcal{H}^l \subset C_b^m(\mathcal{X}) \qquad \|f\|_{C_b^m} \le C_{m,l} \|f\|_{\mathcal{H}^l} \tag{4.5}$$

其中 $C_{m,\ l}$ 是一个仅依赖于 m 和 l的常数。

取 $l = s = \lfloor p/2 \rfloor + 1$ 和 $m = 0$，我们看到

$$C_b^s(\mathcal{X}) \subset \mathcal{H}^s \subset C_b(\mathcal{X}),$$

其中对于某个常数 C_6，$\|f\|_\infty \le C_6\|f\|_{\mathcal{H}^s}$ 满足 $\forall f \in \mathcal{H}^s$。这种范数关系表明，$\mathcal{H}^s$ 是一个具有有界核 $s(\cdot,\cdot)$ 的再生核希尔伯特空间。

4.3. 第二部分：算子差的界

与（4.4）类似，我们定义乘法算子

$$D^{1/2} : \mathcal{H}^s \to \mathcal{H}^s \text{ as } (D^{1/2}f)(x) = \sqrt{d(x)}f(x), \tag{4.6}$$

$$D : \mathcal{H}^s \to \mathcal{H}^s \text{ as } (Df)(x) = d(x)f(x). \tag{4.7}$$

在本小节中，我们证明 $D^{1/2}, D^{-1/2}, D, D_n^{1/2}, D_n^{-1/2}, D_n \in \mathcal{L}(\mathcal{H}^s)$ 及其算子范数均有适当的界，即

引理 4.3. 在一般假设下，以下所有算子都是 $\mathcal{L}(\mathcal{H}^s)$ 中的有界线性算子，并且存在适当的常数 $C_7 > 0$，使得

$$\|D^{1/2}\|_{\mathcal{H}^s\to\mathcal{H}^s}, \|D^{-1/2}\|_{\mathcal{H}^s\to\mathcal{H}^s}, \|D_n^{1/2}\|_{\mathcal{H}^s\to\mathcal{H}^s}, \|D_n^{-1/2}\|_{\mathcal{H}^s\to\mathcal{H}^s} \le C_7 \tag{4.8}$$

$$\|(D^{1/2} + D_n^{1/2})^{-1}\|_{\mathcal{H}^s\to\mathcal{H}^s} \le C_7, \quad \|D - D_n\|_{\mathcal{H}^s\to\mathcal{H}^s} \le C_7\|d - d_n\|_{\mathcal{H}^{p+2}} \tag{4.9}$$

证明。令 $C_8 = \|k\|_{C_b^{p+2}(\mathcal{X}\times\mathcal{X})}$。对于任意 $x \in \mathcal{X}$，显然 $k_x := k(\cdot, x) \in C_b^{p+2}(\mathcal{X})$，其中 $\|k_x\|_{C_b^{p+2}} \le C_8$。由于 d 和 d_n 是 k_x 的某种加权平均，因此

$$\|d\|_{C_b^{p+2}}, \|d_n\|_{C_b^{p+2}} \le C_8.$$

由于 $d\ d_n$ $C_b^{p+2}\ \mathcal{X}$ $\kappa_l\ \kappa_u$ $k\ \cdot\ \cdot$ $d^{1/2}\ d_n^{1/2}\ d^{-1/2}\ d_n^{-1/2} \in$，继承了来自（，）的逐点界，因此我们知道，，，（）满足

$$\|d^{1/2}\|_{C_b^{p+2}}, \|d_n^{1/2}\|_{C_b^{p+2}}, \|d^{-1/2}\|_{C_b^{p+2}}, \|d_n^{-1/2}\|_{C_b^{p+2}} \le C_9.$$

接下来，根据 Burenkov 第 4 章引理 15 [6] 可知，对于 $g \in C_b^s(\mathcal{X})$ 和 $f \in \mathcal{H}^s$，有 $gf \in \mathcal{H}^s$ 且

$$\|gf\|_{\mathcal{H}^s} \le \|g\|_{C_b^s}\|f\|_{\mathcal{H}^s}. \tag{4.10}$$

我们可以利用这个不等式证明 $D^{1/2}$，$D^{-1/2}$，D，$D_n^{1/2}$，$D_n^{-1/2}$，$D_n \in \mathcal{L}(\mathcal{H}^s)$，并界定它们的算子范数。例如，将 $g = d^{1/2}$，$d_n^{1/2}$，$d^{-1/2}$，$d_n^{-1/2}$代入（4.10），并注意到对于这些 g，$\|g\|_{C_b^s} \le \|g\|_{C_b^{p+2}}$，因为$p + 2 > s$，我们得出

$$\|D^{1/2}\|, \|D^{-1/2}\|, \|D_n^{1/2}\|, \|D_n^{-1/2}\| \le C_9.$$

注意到，根据嵌入定理，$\mathcal{H}^{p+2}$可以嵌入到 $C_b^s(\mathcal{X})$中，因此代入 $d - d_n \in C_b^{p+2}(\mathcal{X}) \subset \mathcal{H}^{p+2}$，我们得到

$$\|D - D_n\| \le \|d - d_n\|_{C_b^s} \le C_{10}\|d - d_n\|_{\mathcal{H}^{p+2}}.$$

对于 $\|(D^{1/2}+D_n^{1/2})^{-1}\|$的界，我们基本上遵循相同的思路。首先给出 $\|d^{1/2}+d_n^{1/2}\|_{C_b^{p+2}}$ 的界，然后论证 $d^{1/2}+d_n^{1/2}$具有逐点上下界。于是可得 $\|(d^{1/2}+d_n^{1/2})^{-1}\|_{C_b^{p+2}} \le C_{11}$，并且通过（4.10）可知 $\|(D^{1/2}+D_n^{1/2})^{-1}\|_{\mathcal{H}^s}$, $_{\mathcal{H}^s} \le C_9$。令 C_7 为 C_9 到C_{11} 中的最大值，即可完成证明。

□

引理 4.4. 在一般假设下，我们有

$$\|T-\widehat{T}_n\|_{HS} \le C_{12}\Big((\|K_n\|_{HS}+\|K\|_{HS})\|d-d_n\|_{\mathcal{H}^{p+2}}+\|K-K_n\|_{HS}\Big).$$

证明. 首先注意到

$$\begin{aligned} & D^{-1/2}-D_n^{-1/2} \\ &= D_n^{-1/2}(D_n^{1/2}-D^{1/2})D^{-1/2} \\ &= D_n^{-1/2}(D_n-D)(D_n^{1/2}+D^{1/2})^{-1}D^{-1/2}. \end{aligned}$$

应用引理 4.3中的界，我们得到

$$\|D^{-1/2}-D_n^{-1/2}\| \le C_7^4\|d-d_n\|_{\mathcal{H}^{p+2}}.$$

我们还有如下分解

$$\begin{aligned} & D^{-1/2}KD^{-1/2}-D_n^{-1/2}K_nD_n^{-1/2} \\ &= D^{-1/2}K(D^{-1/2}-D_n^{-1/2})+(D^{-1/2}K-D_n^{-1/2}K_n)D_n^{-1/2} \\ &= D^{-1/2}K(D^{-1/2}-D_n^{-1/2})+D^{-1/2}(K-K_n)D_n^{-1/2}+(D^{-1/2}-D_n^{-1/2})K_nD_n^{-1/2}. \end{aligned}$$

对两边取希尔伯特–施密特范数，得到[12]

$$\begin{aligned} & \|T-\widehat{T}_n\|_{HS} \\ &\le C_7^5\|K\|_{HS}\|d-d_n\|_{\mathcal{H}^{p+2}}+C_7^2\|K-K_n\|_{HS}+C_7^5\|K_n\|_{HS}\|d-d_n\|_{\mathcal{H}^{p+2}} \\ &\le C_{12}\Big((\|K_n\|_{HS}+\|K\|_{HS})\|d-d_n\|_{\mathcal{H}^{p+2}}+\|K-K_n\|_{HS}\Big) \end{aligned}$$

其中 C_{12}经过了适当选取。 □

4.4. 希尔伯特空间上的集中性

在本小节中，我们将证明 K， K_n 都是从 $\mathcal{H}^s$到 $\mathcal{H}^s$ 的 Hilbert-Schmidt 算子，并建立关于$\|d-d_n\|_{\mathcal{H}^{p+2}}$ 和$\|K-K_n\|_{HS}$的一些集中性结果。借助这些结果和引理 4.4，我们将能够界定 $\|T-\widehat{T}_n\|_{HS}$。所需的集中界通过以下关于（复）希尔伯特空间中集中性的定理获得（参见 Rosasco、Belkin 和 Vito 的第 2.4 节 [29]）。

[12]我们尚未证明 K， K_n 是从 H^s 到 H^s 的 Hilbert-Schmidt 算子。这一点将在引理4.6中证明

引理 4.5. 设 ξ_1，...，ξ_n 是取值于可分（复）希尔伯特空间 $\mathcal{H}$ 中的零均值独立随机变量，且对所有 $i \in [n]$ 都有 $\|\xi_i\|_{\mathcal{H}} \leq C$ 。则以至少 $1-2e^{-\tau}$ 的概率，有

$$\|\frac{1}{n}\sum_{i=1}^{n}\xi_i\|_{\mathcal{H}} \leq \frac{C\sqrt{2\tau}}{\sqrt{n}}.$$

利用该引理，我们可以证明

引理 4.6。在一般假设下，以下事实成立：

1. 对于某个常数 C_{13}，以 $1-2e^{-\tau}$ 的置信度

$$\|d-d_n\|_{\mathcal{H}^{p+2}} \leq C_{13}\frac{\sqrt{\tau}}{\sqrt{n}}.$$

2. K 和 K_n 都是从 $\mathcal{H}^s$ 到 $\mathcal{H}^s$ 的 Hilbert-Schmidt 算子，并且存在一个与 n 无关的常数 C_{14} ，使得它们的 Hilbert-Schmidt 范数 $\|K_n\|_{HS}$、$\|K\|_{HS} \leq C_{14}$ 有界。

3. 对于某个常数 C_{15}，以 $1-2e^{-\tau}$ 的置信度

$$\|K-K_n\|_{HS} \leq C\frac{\sqrt{\tau}}{\sqrt{n}}.$$

证明。对于第 1 项，考虑随机变量 $\xi_i = k(\cdot, X_i) - d \in \mathcal{H}^{p+2}$ 当$i \in [n]$时。它们显然是零均值的。根据引理 4.3 的证明，我们有 d，$k(\cdot, X_i) \in C_b^{p+2}(\mathcal{X})$。因此，我们得到

$$\|\xi_i\|_{\mathcal{H}^{p+2}} \leq \|k(\cdot,X_i)\|_{\mathcal{H}^{p+2}} + \|d\|_{\mathcal{H}^{p+2}} \leq C_{\mathcal{X}}\left(\|k(\cdot,X_i)\|_{C_b^{p+2}} + \|d\|_{C_b^{p+2}}\right)$$

其中 $C_{\mathcal{X}}$ 是一个取决于有界集 $\mathcal{X}$的勒贝格测度的常数。这表明 ξ_i是有界的。由于 $\mathcal{H}^{p+2}$ 是一个可分Hilbert空间，我们应用引理 4.5 并得出，以概率 $1-2e^{-\tau}$

$$\|d-d_n\|_{\mathcal{H}^{p+2}} \leq C_{13}\frac{\sqrt{\tau}}{\sqrt{n}}.$$

对于第 2 项，任取 $x \in \mathcal{X}$，并考虑算子 $\langle\cdot, s_x\rangle_{\mathcal{H}^s}k_x$ where$s_{X_i} := s(\cdot, X_i)$。事实上，该算子是一个从 $\mathcal{H}^s$到 $\mathcal{H}^s$的 Hilbert-Schmidt 算子。为此，注意到 $\|\langle\cdot, s_x\rangle_{\mathcal{H}^s}k_x\|_{HS} = \|s_x\|_{\mathcal{H}^s}\|k_x\|_{\mathcal{H}^s}$。运用证明第 1 项时的相同推理，我们可知 $\|k_x\|_{\mathcal{H}^s}$ 在 $\forall x \in \mathcal{X}$上有一致的上界。剩下要证明 $\|s_x\|_{\mathcal{H}^s}$ 有一致的上界。令 $\delta_x : \mathcal{H}^s \to \mathbb{C}$为求值泛函，即 $\delta_x(f) = f(x)$。根据嵌入定理，我们知道对所有 $x \in \mathcal{X}$，都有 $\|\delta_x\|_{op} \leq C_6$。但 s_x 也诱导出这个点求值泛函，因此由里斯表示定理，$\|s_x\|_{\mathcal{H}^s} = \|\delta_x\|_{op} \leq C_6$。于是，对某个 C_{14}，对所有 $x \in \mathcal{X}$，都有 $\|\langle\cdot, s_x\rangle_{\mathcal{H}^s}k_x\|_{HS} \leq C_{14}$。现在令 x为随机变量。我们看到 $\|K\|_{HS} = \|\mathbb{E}\langle\cdot, s_{X_i}\rangle_{\mathcal{H}^s}k_{X_i}\|_{HS} \leq \mathbb{E}\|\langle\cdot, s_{X_i}\rangle_{\mathcal{H}^s}k_{X_i}\|_{HS} \leq C_{14}$，即 K是

希尔伯特-施密特。运用相同的推理，我们可知第 2 项中关于 K_n 的断言也成立。

对于第 3 项，考虑随机变量 $\omega_i := \langle \cdot, s_{X_i} \rangle_{\mathcal{H}^s} k_{X_i} - K$。由第 2 项可知 $\omega_i \in HS(\mathcal{H}^s)$。由于 $\mathcal{H}^s$ 是可分的，希尔伯特空间 $HS(\mathcal{H}^s)$ 也是可分的。我们还知道 ω_i 是零均值的，并且 $\|\omega_i\|_{HS} \leq 2C_{14}$ 是有界的。因此，我们可以应用引理 4.5，从而得到其以概率 $1-2e^{-\tau}$

$$\|K - K_n\|_{HS} \leq C_{15} \frac{\sqrt{\tau}}{\sqrt{n}}$$

对于 $C_{15} := 2C_{14}$。 □

结合引理 4.4和 4.6，我们得到所需结果

命题 4.7。在一般假设下，以概率 $1-4e^{-\tau}$，我们有

$$\|T - \widehat{T}_n\|_{HS} \leq C_{16} \frac{\sqrt{\tau}}{\sqrt{n}}$$

对于某个常数 C_{16}。

证明。 并集界以及直接应用引理 4.4 即可完成证明。 □

4.5. 检查一般理论的条件

在第 3 节的前三段中，我们列出了应用一般理论必须满足的条件。在前面的三个小节中，我们已经隐含地检查了其中的大部分条件，但为保证完备性，我们在此总结并证明所有这些条件。

引理 4.8. 在一般条件下，以下事实成立：

1. Sobolev空间 H^s 是 $L^2(\mathcal{X}, \mathbb{P})$ 的子空间。2. $\mathcal{H}^s$ 范数 $\|\cdot\|_{\mathcal{H}^s}$ 强于无穷范数。3. T、T_n 从 $\mathcal{H}_s$ 到 $\mathcal{H}_s$都是希尔伯特-施密特的。4. T 的所有特征值（按重数计）都可以排列成一个由非负实数组成的递减（可能是无限）序列 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_K > \lambda_{K+1} \geq \ldots \geq 0$，且 λ_K 与 λ_{K+1}之间存在正间隔。5. 可以选取前 K 个特征函数 $\{f_i\}_{i=1}^K \subset \mathcal{H}^s$，使其构成 $L^2(\mathcal{X}, \mathbb{P})$ 中函数的标准正交集。6. T_n 具有一个由不递增的实非负特征值组成的序列。

证明。对于第 1 项，这是因为 H^s 是 $L^2(\mathcal{X}, dx)$ 的子空间，并且根据我们对 $\mathcal{X}$ 和 $\mathbb{P}$ 的假设，$L^2(\mathcal{X}, dx)$ 与 $L^2(\mathcal{X}, \mathbb{P})$ 是同一个空间。首先，由于 $\mathbb{P}$ 的底层 σ-代数是勒贝格 σ-代数，可测函数的集合相同。如果 $f \in L^2(\mathcal{X}, dx)$，那么 f 也属于 $L^2(\mathcal{X}, \mathbb{P})$，因为 $\int_{\mathcal{X}} f \bar{f} dP = \int_{\mathcal{X}} f \bar{f} p(x) dx \leq p_u \int_{\mathcal{X}} f \bar{f} dx < \infty$。反之亦然。不难看出，我们的假设保证勒贝格测度

相对于 $\mathbb{P}$ 绝对连续，其密度为 $1/p(x)$，几乎处处成立。注意到 $1/p(x) < 1/p_l$，我们即可证明反向结论。

第 2 项是 Sobolev嵌入定理的推论，并且在前面的小节中已被反复使用。第 3 项是引理 4.3 和4.6的共同推论。

对于第 4 项和第 5 项，我们先证明 T作为从 $L^2(\mathcal{X},\mathbb{P})$ 到 $L^2(\mathcal{X},\mathbb{P})$ 的算子是正的、自伴的且为希尔伯特-施密特。令 $h(x,y) := k(x,y)/\sqrt{d(x)d(y)}$ 表示归一化核。自伴性源于从 $k(\cdot,\cdot)$ 继承的（共轭）对称性 $h(\cdot,\cdot)$:

$$\langle f, Tg\rangle = \int f(x)\Big(\int \overline{h(x,y)}\,\overline{g(y)}dP(y)\Big)dP(x) = \iint h(y,x)f(x)\overline{g(y)}dP(x)dP(y),$$
$$\langle Tf, g\rangle = \int \Big(h(x,y)f(y)dP(y)\Big)\overline{g(x)}dP(x) = \iint h(x,y)f(y)\overline{g(x)}dP(x)dP(y).$$

因此我们看到 $\langle f, Tg\rangle = \langle Tf, g\rangle$，即 T是自伴的。为了说明 T为何是希尔伯特-施密特，令实值函数 $\{e_i\}_{i=1}^{\infty}$ 构成 $L^2(\mathcal{X},\mathbb{P})$ 的一组标准正交基。我们计算

$$\|T\|_{HS}^2 = \sum_{i=1}^{\infty}\left(\int\Big(\int h(x,y)e_i(y)dP(y)\Big)^2 dP(x)\right) = \int\Big(\sum_{i=1}^{\infty}\langle h_x, e_i\rangle_{L^2}^2\Big)dP(x)$$
$$= \iint h^2(x,y)dP(x)dP(y) \le \kappa_u^2/\kappa_l^2 < \infty.$$

正性部分稍微复杂一些。为了证明 T 是正的，我们需要证明对于 $\forall f \in L^2(\mathcal{X},\mathbb{P})$

$$\langle Tf, f\rangle = \iint k(x,y)\overline{f(x)}f(y)dP(x)dP(y) \ge 0.$$

固定样本量 n，并抽取独立同分布样本 $X_1, X_2, \ldots, X_n \sim \mathbb{P}$。由于核 $k(\cdot,\cdot)$ 是正定的，二次型

$$\frac{1}{n^2}\sum_{i,j=1}^{n} k(X_i,X_j)\overline{f(X_i)}f(X_j)$$

无论抽取什么样本，都是非负的。因此，该二次型的期望也非负。简单计算可知，这个期望实际上为

$$\frac{n-1}{n}\langle Tf, f\rangle + \frac{1}{n}\int k(x,x)\overline{f(x)}f(x)dP(x).$$

由于根据我们的假设，$k(x,x) \le \kappa_u$且 $f \in L^2(\mathcal{X},\mathbb{P})$，可知 $\int k(x,x)f(x)f(x)dP(x)$ 是有限的。由于 n可以任意大，$\langle Tf, f\rangle$必须非负。

根据第 2.2 节介绍的正自伴 Hilbert-Schmidt 算子谱理论，我们可以立即看出，条目

4 和 5 的大部分内容都成立。条目 5 尚需验证的是 $\{f_i\}_{i=1}^K \subset \mathcal{H}^s$，这可由假设 $\{f_i\}_{i=1}^K \subset C_b^{p+2}(\mathcal{X})$ 推出。条目 4 中关于特征值间隙的部分也由一般假设所保证。条目 4 中需要注意的是，其中的特征值和特征向量是在 T 是从 $\mathcal{H}^s$ 到 $\mathcal{H}^s$ 的算子的前提下定义的。但由于 $\mathcal{H}^s$ 是 $L^2(\mathcal{X},\mathbb{P})$ 的子空间，因此，与将 T 视为 $\mathcal{L}(L^2(\mathcal{X},\mathbb{P}))$ 中的算子相比，我们只能得到更少的特征值。此外，由于一般假设蕴含 $\{f_i\}_{i=1}^K \subset \mathcal{H}^s$，将算子从 $\mathcal{L}(L^2(\mathcal{X},\mathbb{P}))$ 限制到 $\mathcal{H}^s$ 后，主特征空间保持不变。

对于第 6 项，这是因为 $\asymp T_n$ 的谱与 L_n 这一对称半正定核矩阵的谱之间存在关系。将其中的 4.1 引理中的 $C_b(\mathcal{X})$ 替换为 $\mathcal{H}^s$ 后，类似的引理也成立。

□

由于引理 4.8，我们可以应用定理 3.2 的略作修改的版本（参见定理 3.2 的证明）来得到以下结论。

命题 4.9：对于某个常数 C_{17}、C_{18}，当 $n > C_{17}\tau$ 时，我们以 $1-4e^{-\tau}$ 的置信度有

$$\|V_1 - (V_1+V_2Y)(I+Y^*Y)^{-1/2}\|_{2\to\infty} \le C_{18}\frac{\sqrt{\tau}}{\sqrt{n}} \tag{4.11}$$

证明。 根据命题 4.7，我们以概率 $1-4e^{-\tau}$ 得到，

$$\|T - \widehat{T}_n\|_{HS} \le C_{16}\frac{\sqrt{\tau}}{\sqrt{n}}.$$

我们可以将 C_{17} 取为充分大，使得 $C_{16}\frac{\sqrt{\tau}}{\sqrt{n}} \le C_{16}/\sqrt{C_{17}} \le C_1$，其中 C_1 是定理 3.2 中的常数。因此，根据定理 3.2 证明中的一个中间步骤，我们得到

$$\|V_1 - (V_1+V_2Y)(I+Y^*Y)^{-1/2}\|_{2\to\infty} \le C_{18}\frac{\sqrt{\tau}}{\sqrt{n}}.$$

□

4.6. 第五部分：由 V_1 引起的误差

我们处理算子 V_1 在 $L^2(\mathcal{X},\mathbb{P})$ 中不具有正交归一列所引入的误差。引入简写 $W := (V_1+V_2Y)(I+Y^*Y)^{-1/2}$，则有

$$\begin{aligned}\|V_1 - \widehat{V}_1Q\|_{2\to\infty} &\le \|V_1 - W\|_{2\to\infty} + \|W - \widehat{V}_1Q\|_{2\to\infty}\\ &= \|V_1 - W\|_{2\to\infty} + \|W - WW^*\widehat{V}_1Q\|_{2\to\infty}\\ &\le \|V_1 - W\|_{2\to\infty} + \|W\|_{2\to\infty}\|Q^T - W^*\widehat{V}_1\|_2. \end{aligned}\tag{4.12}$$

这里，第二步中的等式成立，是因为 V_1 和 W 张成相同的主特征空间[13]，并且 W 在 $L^2(\mathcal{X},\mathbb{P})$ 中具有正交归一列。考察 4.12，我们看到 $\|V_1 - W\|_{2\to\infty}$ 由命题 4.9界定，而 $\|W\|_{2\to\infty}$ 大致也有 $\|V_1\|_{2\to\infty}$ 类似的界，因此问题归结为 W^*V_1 与 $\mathbb{C}^{K\times K}$中的某个酉矩阵相距多“远”。事实上，有

引理 4.10。假设 W^*V_1 的所有奇异值都小于 2，则存在酉矩阵 $Q\in\mathbb{C}^{K\times K}$，使得

$$\|Q - W^*\widehat{V}_1\|_2 \le 2\|W\|^2_{2\to H^s}\sup_{\|g\|_{\mathcal{H}^s}\le 1}\left|P_n g^2 - Pg^2\right|.$$

证明。假设 W^*V_1 具有奇异值分解 $W^*V_1 = A\Sigma B^*$，则有 $\Sigma = A^*W^*V_1B$。令 $g_i := WAe_i$ 为 WA中的第 i 列，$\tilde{g}_i := V_1Be_i$ 为 $\widehat{V}_1B$中的第 i 列，其中 $\{e_i\}_{i=1}^K$ 是 $\mathbb{R}^K$中的标准基。我们知道$\Sigma = (\Sigma_{ij})$，其中 $\Sigma_{ij} = \langle\tilde{g}_i,\ g_j\rangle_{L^2(\mathcal{X},\ \mathbb{P})}$。

由于 A、B是酉矩阵，$\{g_i\}_{i=1}^K$在 $L^2(\mathcal{X},\mathbb{P})$ 中是正交归一的，并且$\{\tilde{g}_i\}_{i=1}^K$在 $L^2(\mathcal{X},\mathbb{P}_n)$ 中也是正交归一的。因此，g_1与 g_2、……、g_K正交。同时，根据 Σ的对角结构，我们知道 $\tilde{g}_1$也与 g_2、……、g_K正交。这表明 g_1 与 $\tilde{g}_1$共线。此外，由于 Σ 的对角元素是实正值，我们知道 $\tilde{g}_1 = g_1/\|g_1\|_{L^2(\mathcal{X},\ \mathbb{P}_n)}$。事实上，这对所有 $i\in[K]$均成立，即 $\tilde{g}_i = g_i/\|g_i\|_{L^2(\mathcal{X},\ \mathbb{P}_n)}$。取 $Q = AB^*$，它是一个酉矩阵，有

$$\|Q - W^*\widehat{V}_1\|_2 = \|\Sigma - I\|_2 = \max_{i\in[K]}|1 - \langle\tilde{g}_i, g_i\rangle| = \max_{i\in[K]}|1 - \frac{1}{\|g_i\|_{L^2(\mathcal{X},\mathbb{P}_n)}}|. \quad (4.13)$$

根据我们对奇异值的假设，我们知道对于所有 $i\in[K]$、$\frac{1}{\|g_i\|_{L^2(\mathcal{X},\mathbb{P}_n)}} = \langle\tilde{g}_i, g_i\rangle \le 2$。注意，对于 $x\ge\frac{1}{2}$、$|1-\frac{1}{x}| \le 2|x-1| \le 2|x^2-1|$，我们看到

$$\max_{i\in[K]}|1-\frac{1}{\|g_i\|_{L^2(\mathcal{X},\mathbb{P}_n)}}| \le 2\max_{i\in[K]}|1-\|g_i\|^2_{L^2(\mathcal{X},\mathbb{P}_n)}| = 2\max_{i\in[K]}\left|\frac{1}{n}\sum_{j=1}^n|g_i(X_j)|^2 - \mathbb{E}|g_i(X)|^2\right|. \quad (4.14)$$

由于 g_i 依赖于样本 $\{X_i\}_{i=1}^n$，它们是随机的。它们的共同点是具有有界的 $\mathcal{H}^s$ 范数，这是因为

$$\|g_i\|_{\mathcal{H}^s} = \|WAe_i\|_{\mathcal{H}} \le \|W\|_{2\to\mathcal{H}^s}\|A\|_2\|e_i\|_2 = \|W\|_{2\to\mathcal{H}^s}.$$

因此，

$$\max_{i\in[K]}\left|\frac{1}{n}\sum_{j=1}^n|g_i(X_j)|^2 - \mathbb{E}|g_i(X)|^2\right| \le \sup_{\|g\|_{\mathcal{H}^s}\le\|W\|_{2\to\mathcal{H}^s}}\left|P_n|g|^2 - P|g|^2\right| = \|W\|^2_{2\to\mathcal{H}^s}\sup_{\|g\|_{\mathcal{H}^s}\le 1}\left|P_n|g|^2 - P|g|^2\right|. \quad (4.15)$$

Chaining (4.15) and (4.14) completes the proof. □

[13]由于 $\widehat{V}_1$ 是由 L_n 的特征向量构造的，而这些特征向量线性无关，因此这些列不可能是 $\mathcal{H}^s\subset C_b(\mathcal{X})$ 中的线性相关函数。

利用Dudley不等式以及Sobolev空间中关于覆盖数的标准结果，我们可以证明（证明见附录）

引理4.11。对于我们选择的 s，我们有概率为 $1-4\exp(-\tau)$

$$\sup_{\|g\|_{\mathcal{H}^s}\le 1}\left|P_n|g|^2-P|g|^2\right|\le\frac{C_{19}+C_{20}\sqrt{\tau}}{\sqrt{n}}.$$

现在我们可以证明定理 4.2。

定理 *4.2* 的证明。对于固定的样本量 n，令 $\mathcal{E}_{n,1}$ 表示命题 4.7 中的集中性成立的事件，令 $\mathcal{E}_{n,2}$ 表示引理 4.11中的集中性成立的事件。从现在起，我们以交集 $\mathcal{E}_{n,1}\cap\mathcal{E}_{n,2}$为条件，该交集发生的概率大于或等于 $1-8e^{-\tau}$。

首先，在此事件下，我们知道命题4.9 也成立。因此我们有

$$\|V_1-W\|_{2\to\infty}\le C_{18}\frac{\sqrt{\tau}}{\sqrt{n}}.$$

因此 $\|V_1\|_{2\to\infty}$ 接近 $\|W\|_{2\to\infty}$。由于定理 4.2, $n/\tau\ge C_4$，并且我们可以自由选择 C_5，因此可以令 C_5 足够大，使得 $\|W\|_{2\to\infty}\le 2\|V_1\|_{2\to\infty}$。仿照命题 4.9 的证明，我们可以类似地证明 $\|V_1-W\|_{2\to\mathcal{H}^s}$的量级为 $\sqrt{\tau}/\sqrt{n}$。因此可以假设 C_4也足够大，从而确保 $\|W\|_{2\to\mathcal{H}^s}\le 2\|V_1\|_{2\to\mathcal{H}^s}$。

同时，由于引理 4.11 中的大数定律具有一致性，我们总可以通过将 C_4 设得足够大，使得$\|g_i\|_{L^2(\mathcal{X},\mathbb{P}_n)}$ 大于 $1/2$ 。因此，引理 4.10 中关于奇异值的条件得以满足，由此可知存在酉矩阵 Q 使得

$$\|Q-W^*\widehat{V}_1\|_2\le 2\|W\|_{2\to H^s}^2\frac{C_{19}+C_{20}\sqrt{\tau}}{\sqrt{n}}\le 2\|W\|_{2\to H^s}^2\frac{(C_{19}+C_{20})\sqrt{\tau}}{\sqrt{n}},$$

其中我们假设 $\tau\ge 1$。由于对于定理 4.2 这类集中性结果而言， τ 本来就足够大，因此这一假设并无实质影响。

回到（4.12），我们有

$$\begin{aligned}\|V_1-\widehat{V}_1Q\|_{2\to\infty}&\le\|V_1-W\|_{2\to\infty}+\|W\|_{2\to\infty}\|Q^T-W^*\widehat{V}_1\|_2 &(4.16)\\&\le C_{18}\frac{\sqrt{\tau}}{\sqrt{n}}+2\|W\|_{2\to\infty}\|W\|_{2\to H^s}^2\frac{(C_{19}+C_{20})\sqrt{\tau}}{\sqrt{n}} &(4.17)\\&\le C_{18}\frac{\sqrt{\tau}}{\sqrt{n}}+16\|V_1\|_{2\to\infty}\|V_1\|_{2\to H^s}^2\frac{(C_{19}+C_{20})\sqrt{\tau}}{\sqrt{n}}. &(4.18)\end{aligned}$$

设置 $C_3=C_{18}+16\|V_1\|_{2\to\infty}\|V_1\|_{2\to H^s}^2(C_{19}+C_{20})$，从而完成证明。

□

5. 讨论

我们首先想评论一下定理3.2 与谱投影的集中性之间的关系（参见 Rosasco、Belkin 和 Vito [29]中的命题 22）。事实上，我们的结果很容易推出谱投影的集中性。为说明这一点，只需注意到投影之差可以写成 $\widetilde{\ }$ $\widetilde{\ }$如下形式：$V_1V_1^* - V_1V_1^*$然后应用三角不等式。我们相信，可以从希尔伯特空间中谱投影的集中性 $\mathcal{H}$ 推导出定理 3.2，但过程十分棘手。从较高层面来看，我们需要将扰动算子的主不变子空间的一组正交归一基投影到未扰动算子的主不变子空间上，然后对这些投影进行格拉姆-施密特正交化。在此过程中，我们需要在 $\mathcal{H}$ 与 L^2（$\mathcal{X}$，$\mathbb{P}$）之间来回转换，并且预计会遇到无数琐碎而棘手的技术细节。但我们相信，在 $\mathcal{H}$ 诱导算子范数下，谱投影的集中性等价于定理3.2。

我们还想谈谈牛顿-坎托罗维奇定理的普适性。也就是说，算子方程 (3.5) 不必局限于 $\mathcal{L}(\mathbb{C}^K, \mathcal{H})$ 的空间。我们可以对 (3.5) 稍作修改，使其涉及 $L^2(\mathcal{X}, \mathbb{P})$、$C_b(\mathcal{X})$ 或 $C_b^1(\mathcal{X})$，从而诱导出一个不变子空间，仍然可以应用牛顿-坎托罗维奇定理来求解。例如，我们应该可以将本文中的每个 $\mathcal{H}$ 都替换为 $C_b^1(\mathcal{X})$，并相应地重写证明，使整个论证依然成立。需要谨慎的是，要使样本层面算子收敛到总体算子（按算子范数），所使用的函数空间必须具有某种“光滑性”。再生核希尔伯特空间所具有的光滑性或 $C_b^1(\mathcal{X})$ 所具有的光滑性都可以；但像 $C_b(\mathcal{X})$ 或 $L^2(\mathcal{X}, \mathbb{P})$ 这样的空间就不行，因为其中的函数即使范数很小，也可能剧烈振荡，从而可以构造对抗性函数破坏算子范数收敛。von Luxburg、Belkin 和 Bousquet [41] 也提到了这一点。

最后，我们想对复值函数这一假设，以及定理 3.2需要酉矩阵这一事实 Q 发表评论。我们认为，既然所有内容都是实数，酉矩阵只是一个人为产物而非必要条件，因此我们的证明应该可以修改为只需正交矩阵（尽管我们目前还不知道如何做到这一点）。我们还确认，对于几乎所有引理和定理，除了定理3.8之外，我们都可以使用实Hilbert空间或巴拿赫空间以及实值函数来处理。但从积极的一面来看，使用复数使我们的结果更具一般性，并且可以让我们自由地使用复值核函数，尽管在统计学或机器学习中很少有人利用这种自由。最后但同样重要的是，我们想指出，由于篇幅限制，我们只给出了一个应用，即归一化谱聚类，但我们的一般理论还可以应用于其他问题。例如，可以为核主成分分析得到一致性结果，而且其证明比归一化谱聚类的证明简单得多。我们在附录中给出了这些结果。

参考文献

[1] Abbe, E., Fan, J., Wang, K. 和 Zhong, Y. （2020）。低期望秩随机矩阵的逐元素特征向量分析。《统计年鉴》48 1452。[2] Anselone, P. M. （1971）。集体紧算子逼近理论及其在积分方程中的应用。《自动计算》。普伦蒂斯霍尔。[3] Belkin, M. 和 Niyogi, P. （2003）。用于降维和数据表示的拉普拉斯特征映射。《神经计算》15 1373-1396。[4] Bengio, Y., Paiement, J.-F., Vincent, P., Delalleau, O.,Roux, N. L. 和 Ouimet, M. （2003）。LLE、Isomap、MDS、特征映射和谱聚类的样本外扩展。载于神经信息处理系统第16届国际会议论文集。NIPS'03 177–184。麻省理工学院出版社，剑桥，美国马萨诸塞州。[5] Blanchard, G., Bousquet, O. 和 Zwald, L. （2007）。核主成分分析的统计性质。《机器学习》66 259–294。[6] Burenkov, V. I. （1998）。定义域上的索伯列夫空间。Teubner数学丛书。莱比锡B. G. Teubner出版社。[7] Cape, J., Tang, M. 和 Priebe, C. E. （2019a）。二到无穷范数与奇异子空间几何及其在高维统计中的应用。《统计年鉴》47 2405–2439。[8] Cape, J., Tang, M. 和 Priebe, C. E. （2019b）。信号加噪声矩阵模型：特征向量偏差与波动。《生物统计学》106 243–250。[9] Ciarlet, P. G. （2013）。线性与非线性泛函分析及其应用。工业与应用数学学会，费城，美国宾夕法尼亚州。[10] Coifman, R. R., Lafon, S., Lee, A. B., Maggioni, M., Nadler, B.,Warner, F. 和 Zucker, S. W. （2005）。几何扩散作为调和分析和数据结构定义的工具：扩散映射。《美国国家科学院院刊》102 7426-7431。[11] Cucker, F. 和 Smale, S. （2002）。学习的数学基础。《美国数学学会通报》39 1–49。[12] Damle, A. 和 Sun, Y. （2020）。不变子空间扰动的一致界。SIAM矩阵分析与应用杂志41 1208–1236。[13] Donetti, L. 和 Muñoz, M. A. （2004）。检测网络社区：一种新的系统高效算法。《统计力学杂志：理论与实验》2004 P10012。[14] Donoho, D. L. 和 Grimes, C. （2003）。Hessian特征映射：用于高维数据的局部线性嵌入技术。《美国国家科学院院刊》100 5591-5596。[15] Edmunds, D. E. 和 Triebel, H. （1996）。函数空间、熵数与微分算子。剑桥数学论丛。剑桥大学出版社。

[16] Fan, J., Wang, W. 和 Zhong, Y. （2018）。特征向量扰动界及其在稳健协方差估计中的应用。《机器学习研究杂志》18 1–42。[17] Higgs, B. W., Weller, J. 和 Solka, J. L. （2006）。谱嵌入在图像和微阵列数据中发现有意义（相关）的结构。《BMC生物信息学》7 74。[18] Hoffmann, H. （2007）。用于新颖性检测的核主成分分析。《模式识别》40 863 - 874。[19] Karow, M. 和 Kressner, D. （2014）。关于矩阵不变子空间的扰动界。《SIAM矩阵分析与应用杂志》35 599-618。[20] Kato, T. （1995）。《线性算子扰动理论》。数学经典。施普林格出版社柏林海德堡。[21] Koltchinskii, V. I. （1998）。逼近积分算子的某些随机矩阵的谱投影渐近性。载于《高维概率》（E. Eberlein, M. Hahn 和 M. Talagrand, 编）191–227。Birkh¨auser Basel，巴塞尔。[22] Koltchinskii, V. 和 Giné, E. （2000）。积分算子谱的随机矩阵近似。Bernoulli 6 113–167。[23] Lei, J. 和 Rinaldo, A. （2015）。随机块模型中的谱聚类的一致性。《统计年鉴》43 215–237。[24] Mao, X., Sarkar, P. 和 Chakrabarti, D. （2021）。利用精确的特征向量偏差估计混合成员关系。《美国统计学会杂志》116 1928–1940。[25] Mendelson, S. 和 Pajor, A. （2005）。使用随机向量的椭球近似。载于《学习理论》（P. Auer 和 R. Meir，编）429–443。Springer Berlin Heidelberg，柏林，海德堡。[26] Mendelson, S. 和 Pajor, A. （2006）。关于具有独立行的矩阵的奇异值。Bernoulli 12 761–773。[27] Ng, A. Y., Jordan, M. I. 和 Weiss, Y. （2001）。关于谱聚类：分析与算法。载于《神经信息处理系统进展》849–856。麻省理工学院出版社。[28] Rohe, K., Chatterjee, S. 和 Yu, B. （2011）。谱聚类与高维随机块模型。《统计年鉴》39 1878–1915。[29] Rosasco, L., Belkin, M. 和 Vito, E. D. （2010）。利用积分算子学习。《机器学习研究杂志》11 905-934。[30] Schiebinger, G., Wainwright, M. J. 和 Yu, B. （2015）。核化谱聚类的几何。《统计年鉴》43 819-846。[31] Shawe-Taylor, J., Williams, C. K. I., Cristianini, N. 和 Kan-dola, J. （2005）。关于Gram矩阵的特征谱与核主成分分析的泛化误差。IEEE信息论汇刊 51 2510-2522。[32] Shi, T., Belkin, M. 和 Yu, B. （2009）。数据谱分析：卷积算子的特征子空间与聚类。《统计年鉴》37 3960–3984。[33] Shi, J. 和 Malik, J. （2000）。归一化割与图像分割。IEEE模式分析与机器智能汇刊 22 888-

905。[34] Steinwart, I. 和 Christmann, A. （2008）。支持向量机。Springer Science & Business Media。[35] Stewart, G. （1971）。闭线性算子近似不变子空间的误差界。SIAM 数值分析杂志 8 796-808。
[36] Tenenbaum, J. B. （2000）。非线性降维的全局几何框架。《科学》290 2319-2323。[37] Trillos, N. G., Hoffmann, F. 和 Hosseini, B. （2019）。图拉普拉斯嵌入的几何结构。arXiv:1901.10651。[38] Trillos, N. G. 和 Slepčev, D. （2015）。谱聚类一致性的变分方法。arXiv:1508.01928。[39] Trillos, N. G., Gerlach, M., Hein, M. 和 Slepcev, D. （2018）。随机几何图上图拉普拉斯算子朝向拉普拉斯-贝尔特拉米算子谱收敛的误差估计。arXiv:1801.10108。[40] Vershynin, R. （2018）。《高维概率：数据科学中的导论与应用》。剑桥统计与概率数学系列。剑桥大学出版社。[41] von Luxburg, U., Belkin, M. 和 Bousquet, O. （2008）。谱聚类的一致性。《统计年鉴》36 555-586。[42] Zwald, L. 和 Blanchard, G. （2006）。核主成分分析中特征子空间的收敛。载于《神经信息处理系统进展》18（Y. Weiss、B. Sch¨olkopf 和 J. C. Platt，编）1649–1656。麻省理工学院出版社。

附录 A：证明

A.1. 引理 *3.1*的证明

引理 *3.1*的证明。 对于第 1 项，我们证明双重包含关系： $V_2^{-1}\mathcal{H} \subset V_2^*\mathcal{H}$ 和 $V_2^*\mathcal{H} \subset V_2^{-1}\mathcal{H}$。首先，对于任意 $l \in V_2^{-1}\mathcal{H}$，我们知道 $V_2 l \in \mathcal{H}$，因此 $V_2^*(V_2 l) \in V_2^*\mathcal{H}$。由于 $l = V_2^* V_2 l$，我们得到 $V_2^{-1}\mathcal{H} \subset V_2^*\mathcal{H}$。其次，对于任意 $l \in V_2^*\mathcal{H}$，不失一般性，设 $l = V_2^* h$ 对于某个 $h \in \mathcal{H}$。注意 $V_2 V_2^* h = h - V_1 V_1^* h$。由于 f_1，……，$f_K \in \mathcal{H}$ 根据假设，我们知道 $V_1 V_1^* h \in \mathcal{H}$，因此 $V_2 l = h - V_1 V_1^* h \in \mathcal{H}$。这就证明了另一个包含关系。

对于第 2 项，由于 $\mathcal{H}$ 是 $L^2(\mathcal{X}, \mathbb{P})$ 的子空间，$V_2^*\mathcal{H}$ 是 $V_2^* L^2(\mathcal{X}, \mathbb{P})$ 的子空间，而后者等于 l^2。验证 $(\cdot, \cdot)$ 是一个内积是常规的。下面我们证明 $V_2^{-1}\mathcal{H}$ 的完备性。设 $\{b_i\}_{i=1}^{\infty}$ 是 $V_2^{-1}\mathcal{H}$ 中的一个柯西序列。根据定义 $\|b\|_{V_2^{-1}\mathcal{H}} = \|V_2 b\|_{\mathcal{H}}$，序列 $\{V_2 b_i\}_{i=1}^{\infty}$ 在 $\mathcal{H}$ 中是柯西序列。假设对于某个 $y \in \mathcal{H}$ 有 $V_2 b_i \xrightarrow{\mathcal{H}} y$。由于根据假设 $\|\cdot\|_{\mathcal{H}}$ 范数强于 $\|\cdot\|_{\infty}$，而后者又强于 $\|\cdot\|_{L^2}$，我们知道 $V_2 b_i \xrightarrow{L^2} y$。由于 V_2 的值域在 $L^2(\mathcal{X}, \mathbb{P})$ 中是闭的，我们知道 y 也在其值域中，并且 $y = V_2 V_2^* y$。注意到 $\|b_i - V_2^* y\|_{V_2^{-1}\mathcal{H}} = \|V_2 b_i - V_2 V_2^* y\|_{\mathcal{H}}$，且由于 $V_2 b_i \xrightarrow{\mathcal{H}} y$，右端趋于零，因此我们看到 $V_2^{-1}\mathcal{H}$ 的空间确实是完备的。

对于第 3 项，对于任意满足 $\|\alpha\|_2 = 1$ 的 $\alpha \in \mathbb{C}^K$，我们有

$$\|V_1\alpha\|_{\mathcal{H}} = \|\sum_{i=1}^{K} \alpha_i f_i\|_{\mathcal{H}} \leq \sum_{i=1}^{K} |\alpha_i| \|f_i\|_{\mathcal{H}} \leq \sqrt{K} \max_{i\in[K]} \|f_i\|_{\mathcal{H}}.$$

这表明 $\|V_1\|_{2\to\mathcal{H}} \leq \sqrt{K}$ 最大值$_{i\in[K]} \|f_i\|_{\mathcal{H}}$。正如第 2 项的证明中所指出的，$\|\cdot\|_{\mathcal{H}}$ 范数强于 $\|\cdot\|_{L^2}$ 范数，即 $\|h\|_{L^2} \leq C_{\mathcal{H}}\|h\|_{\mathcal{H}}$ 对于某个常数 $C_{\mathcal{H}}$ 和 $\forall h \in \mathcal{H}$。因此

$$\|V_1^* h\|_2^2 = \sum_{i=1}^{K} \langle f_i, h\rangle_{L^2}^2 \leq \sum_{i=1}^{K} \|f_i\|_{L^2}^2 \|h\|_{L^2}^2 \leq C_{\mathcal{H}}^2 K \|h\|_{\mathcal{H}}^2.$$

由此可见 $\|V_1^*\|_{\mathcal{H}\to 2} \leq C_{\mathcal{H}}\sqrt{K}$。事实上，$\|V_2\|_{V_2^{-1}\mathcal{H}\to\mathcal{H}} = 1$ 是 $\|b\|_{V_2^{-1}\mathcal{H}} = \|V_2 b\|_{\mathcal{H}}$ 在 $\forall b \in V_2^{-1}\mathcal{H}$ 下的直接推论。最后，对于 $\forall h \in \mathcal{H}$，我们有

$$\|V_2^* h\|_{V_2^{-1}\mathcal{H}} = \|V_2 V_2^* h\|_{\mathcal{H}} = \|\left(I - V_1 V_1^*\right) h\|_{\mathcal{H}} \leq \left(1 + \|V_1\|_{2\to\mathcal{H}}\|V_1^*\|_{\mathcal{H}\to 2}\right) \|h\|_{\mathcal{H}}.$$

由此可得 $\|V_2^*\|_{\mathcal{H}\to V_2^{-1}\mathcal{H}} \leq 1 + C_{\mathcal{H}} K$ 最大值$_{i\in[K]} \|f_i\|_{\mathcal{H}}$。 □

A.2. 用于证明定理 3.5 的引理证明

在定理3.5的证明中，我们使用了以下引理。现在陈述并证明该引理。

引理 A.1. 定义为 $\mathcal{T}$ 的算子 $\mathcal{T}\colon\ \mathcal{E} \mapsto \mathcal{E}$ 为 $\mathcal{T}(Y)\colon = T_{22}Y - YT_{11}$ 是单射且满射。此外，$\inf_{\|Y\|_{HS}=1} \left\|T_{22}Y - YT_{11}\right\|_{HS} > 0$。

证明。首先，我们注意到

$$\begin{aligned} T_{11}\colon \mathbb{R}^K &\longrightarrow \mathbb{R}^K \\ (a_1, a_2, \ldots, a_K) &\longmapsto (\lambda_1 a_1, \lambda_2 a_2, \ldots, \lambda_K a_K) \end{aligned}$$

为了证明 $\mathcal{T}$ 是单射且满射，只需证明对于任意 $g \in \mathcal{E}$，存在唯一的 $y \in \mathcal{E}$，使得 $g = \mathcal{T}(y)$。将 $\mathbb{R}^K$ 中的标准正交基记为 $\{e_i\}_{i=1}^K$。由于 T_{11} 具有对角结构，我们看到 $(\mathcal{T}y)(e_i)$ $=(T_{22} - \lambda_i I)ye_i$。因此，为了证明存在唯一的 $y \in \mathcal{E}$ 使得$g = \mathcal{T}(y)$，只需证明对于 $\forall i \in [K]$，存在唯一的 ye_i 使得 $(T_{22} - \lambda_i I)$ $ye_i = ge_i$。为此，只需证明 λ_i 属于 T_{22} 的预解集。这确实成立，因为：1) T_{22} 是从 $\tilde{l}^2$ 到 $\tilde{l}^2$ 的紧算子；2) $\sigma(T_{22}) \subset \{\lambda_{K+1}, \lambda_{K+2}, \ldots\} \cup \{0\}$。第二点显然成立，而第一点由

$$\|T_{22}\|_{HS} = \|F_\perp^* T F_\perp\|_{HS} \leq \|F_\perp^*\|_{op}\|T\|_{HS}\|F_\perp\|_{op} \leq \|T\|_{HS}.$$

接下来，我们证明 $\mathcal{T}$是一个有界算子。一旦 $\mathcal{T}$有界，由于 $\mathcal{T}$既是单射又是满射，且 $\mathcal{E}$ 是一个巴拿赫空间，根据有界逆定理，我们知道 $\mathcal{T}^{-1} \in \mathcal{L}(\mathcal{E})$，这等价于 $\inf_{\|Y\|_{HS}=1} \left\|T_{22}Y - YT_{11}\right\|_{HS} > 0$。

算子 $\mathcal{T}$ 确实是有界的，因为

$$\left\|T_{22}Y - YT_{11}\right\|_{HS} \leq \|T_{22}\|_{HS}\|Y\|_{HS} + \|T_{11}\|_{HS}\|Y\|_{HS} \leq 2\|T\|_{HS}\|Y\|_{HS}.$$

□

我们指出，当 T_{22} 是自伴算子时，这个引理的证明将大大简化。事实上，我们有

$$\left\|T_{22}Y - YT_{11}\right\|_{HS} \geq \left\|YT_{11}\right\|_{HS} - \left\|T_{22}Y\right\|_{HS} \geq \lambda_K\left\|Y\right\|_{HS} - \left\|T_{22}\right\|_{op}\left\|Y\right\|_{HS}$$

由于 T_{22} 是自伴算子，其算子范数就是它的最大特征值，即 λ_{K+1}。在这种情况下，我们立即得到 $\inf_{\|Y\|_{HS}=1}\left\|T_{22}Y-YT_{11}\right\|_{HS} \geq \lambda_K-\lambda_{K+1}$，从而恢复了特征值间隙。在非归一化谱聚类中，我们将 $\mathcal{H}$ 设为与核函数相关联的再生核希尔伯特空间，并声称 T_{22} 是自伴算子。

A.3. 引理*4.11*的证明

我们看到，引理4.11的核心是关于 $\mathcal{H}^s$中单位球的一致大数定律。证明中需要以下两个引理：第一个来自 Cucker和Smale [11] （命题6），第二个来自 Vershynin [40]（定理8.1.6）。

引理 A.2。记为 $\mathcal{G} = \{g \,|\, g \in \mathcal{H}^s, \|g\|_{\mathcal{H}^s} \leq 1\}$。当 $s > p/2$ 时，对于所有$\epsilon > 0$,

$$\log \mathcal{N}(\mathcal{G}, \|\cdot\|_\infty, \epsilon) \leq \left(\frac{C}{\epsilon}\right)^{p/s} + 1$$

对于某个常数 C。

引理 A.3。设 $(X_t)_{t\in T}$ 是度量空间 (T, d) 上具有次高斯增量的随机过程，即

$$\|X_t - X_s\|_{\Psi_2} \leq Kd(t, s) \text{ for all } t, s \in T.$$

那么，对于每个 $u \geq 0$，事件

$$\sup_{t\in T} X_t \leq CK\Big[\int_0^\infty \sqrt{\log \mathcal{N}(T, d, \epsilon)}d\epsilon + u \cdot diam(T)\Big]$$

以至少 $1 - 2exp(-u^2)$ 的概率成立。

现在证明引理 4.11。对于不熟悉以下论证的读者，我们建议参阅 Vershynin [40] 中定理8.2.3的证明。

引理的证明 *4.11*。 我们首先证明，在 $\mathcal{G} = \{g \,|\, g \in \mathcal{H}^s, \|g\|_{\mathcal{H}^s} \leq 1\}$ 上，随机过程 $P_n|g|^2 - P|g|^2$ 具有次高斯增量。对于固定的 f， $g \in \mathcal{G}$，我们有

$$\|P_nf\bar{f}-Pf\bar{f}-P_ng\bar{g}+Pg\bar{g}\|_{\Psi_2} = \frac{1}{n}\|\sum_{i=1}^n Z_i\|_{\Psi_2} \text{ where } Z_i = (f\bar{f}-g\bar{g})(X_i)-\mathbb{E}(f\bar{f}-g\bar{g})(X).$$

因此， Z_i彼此独立且均值为零。由此可得

$$\|P_n f\bar{f} - Pf\bar{f} - P_n g\bar{g} + Pg\bar{g}\|_{\Psi_2} \leq \frac{C_{21}}{n}\Big(\sum_{i=1}^{n}\|Z_i\|_{\Psi_2}^2\Big)^{1/2}.$$

根据中心化引理，我们知道

$$\|Z_i\|_{\Psi_2} \leq C_{22}\|f(X_i)\overline{f(X_i)} - g(X_i)\overline{g(X_i)}\|_{\Psi_2}.$$

注意，由于嵌入，我们有

$$\|f\bar{f}-g\bar{g}\|_\infty \leq \|f(\bar{f}-\bar{g})\|_\infty+\|(f-g)\bar{g}\|_\infty \leq \|f\|_\infty\|f-g\|_\infty+\|g\|_\infty\|f-g\|_\infty \leq 2C_6\|f-g\|_\infty.$$

因此，随机变量 $f(X_i)\overline{f(X_i)} - g(X_i)\overline{g(X_i)}$ 是有界的。由于有界随机变量具有有界的 Ψ_2 范数，我们看到

$$\|f(X_i)\overline{f(X_i)} - g(X_i)\overline{g(X_i)}\|_{\Psi_2} \leq 2C_6C_{23}\|f-g\|_\infty.$$

综合起来，我们有

$$\|P_n f^2 - Pf^2 - P_n g^2 + Pg^2\|_{\Psi_2} \leq \frac{C_{24}}{\sqrt{n}}\|f-g\|_\infty.$$

接下来，容易验证 $\mathrm{diam}(\mathcal{G}) \leq 2C_6$ 和 $\int_0^{2C_6}\sqrt{\log \mathcal{N}(T,d,\epsilon)}d\epsilon < \infty$ (由于我们对 $s = \lfloor p/2\rfloor + 1$ 的选择以及引理 A.2)。因此，事件

$$\sup_{g\in\mathcal{G}} P_n g^2 - Pg^2 \leq \frac{C_{19}+C_{20}u}{\sqrt{n}}$$

以 $1-2exp(-u^2)$ 的概率成立。

用完全相同的论证，我们还可以证明事件

$$\sup_{g\in\mathcal{G}} Pg^2 - P_n g^2 \leq \frac{C_{19}+C_{20}u}{\sqrt{n}}$$

以概率 $1-2exp(-u^2)$ 成立。取并集界，可得

$$\sup_{g\in\mathcal{G}} \left|Pg^2 - P_n g^2\right| \leq \frac{C_{19}+C_{20}u}{\sqrt{n}}$$

以概率 $1-4exp(-u^2)$ 成立。将 u^2改写为 τ，证明即告完备。

□

附录 B：核主成分分析的应用

为进一步展示一般理论的用法，本节将其应用于核主成分分析（核 PCA）。在核 PCA 中，我们从一个度量空间 $\mathcal{X}$、其上的概率测度 $\mathbb{P}$ $\mathcal{X}$，以及一个连续的正定核函数 k： $\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ 出发。在观察到样本

X_1、……、 X_n 独立同分布$\sim \mathbb{P}$ 后，我们关注由它们两两相似性构成的矩阵 $K_n \in \mathbb{R}^{n\times n}$： $K_n = \left[\frac{1}{n}k(X_i, X_j)\right]_{i,\ j=1}^{n}$，并假设将数据映射到特征空间后是中心化的。由于 K_n 是对称且半正定的，因此它具有一个特征值分解。我们将特征对记为（λ_k， v_k），并按特征值降序排列：

$$\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_n \geq 0.$$

特征向量 v_k 经过归一化，使其满足 $\|v_k\|_2 = \sqrt{n}$。然后，矩阵 $V = [v_1, \cdots, v_K] \in \mathbb{R}^{n\times K}$ 由前 K 个主成分构成。

设 $\mathcal{H}$ 为与核 $k(\cdot,\cdot)$ 相关联的再生核希尔伯特空间。回顾一下， $a, b \in \mathcal{H}$ 的张量积，

$$\begin{aligned} a \otimes b : \mathcal{H} &\to \mathcal{H} \\ f &\mapsto \langle b, f\rangle_{\mathcal{H}} a, \end{aligned}$$

是一个线性算子。那么， K_n 对应的算子形式是经验协方差算子

$$\Sigma_n = \frac{1}{n}\sum_{i=1}^{n} k_{X_i} \otimes k_{X_i}, \tag{B.1}$$

其中， $k_{X_i} = k(\cdot, X_i)$ 是在特征映射 $x \mapsto k(\cdot, x)$ 下， $X_i (i = 1, \cdots, n)$ 在 $\mathcal{H}$ 中对应的特征。

K_n 和 Σ_n 的特征值与特征向量密切相关。为表述这种关系，我们定义限制算子ζ: $\mathcal{H} \to \mathbb{R}^n$，其定义为 $\zeta f = \frac{1}{\sqrt{n}}(f(X_1)\ \cdots, f(X_n))^T$。可以验证， ζ的伴随算子 ζ^*: $\mathbb{R}^n \mapsto \mathcal{H}$ 可表示为 $\zeta^*\alpha = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\alpha_i k_{X_i}$，其中 $\alpha = (\alpha_1, \cdots, \alpha_n)^T$。 Σ_n 和 K_n 的特征值与特征向量（函数）在以下意义上密切相关。

引理 B.1. 以下事实成立：

1. $K_n = \zeta\zeta^*$ 和 $\Sigma_n = \zeta^*\zeta$；
2. 如果 (λ, f) 是 Σ_n 的非平凡特征对（即 $\lambda \neq 0$），那么 $(\lambda, \zeta f)$ 是 K_n 的特征对。
3. 如果 (λ, v) 是 K_n 的非平凡特征对，那么 (λ, f)，其中

$$\widehat{f}(x) = \frac{1}{\widehat{\lambda} n}\zeta^{\star} v = \frac{1}{\widehat{\lambda} n}\sum_{i=1}^{n} k(x, X_i) v_i \tag{B.2}$$

是 Σ_n 的特征对，且 $f \in \mathcal{H}$。此外，对 f 的这一选择满足 $\|f\|_{L^2(\mathcal{X},\ \mathbb{P}_n)} = 1$，并且 f 在样本点上的限制与 v 一致，即 $\zeta f = v$。

证明。对于第一项，很容易验证 $\zeta\zeta^*$： $\mathbb{R}^n \mapsto \mathbb{R}^n$ 是由 K_n 定义的线性变换。至于另一半，注意到

$$\zeta^*\zeta f = \frac{1}{n}\sum_{i=1}^{n} f(X_i) k(x, X_i).$$

同时，根据再生性质

$$\Sigma_n f = \frac{1}{n}\sum_{i=1}^{n}\langle k_{X_i}, f\rangle_{\mathcal{H}} k_{X_i} = \frac{1}{n}\sum_{i=1}^{n} f(X_i)k(x, X_i).$$

因此，我们得出二者相等。

对于第二项，由于根据假设 $\zeta^*\zeta f = \lambda f$，我们有 $K_n\zeta f = \zeta\zeta^*\zeta f = \lambda\zeta f$，这正是该陈述所表明的内容。

对于第三项，如果 $(\widehat{\lambda}, v)$ 是 K_n 的一个特征对，我们验证 $(\widehat{\lambda}, \widehat{f})$ 是 Σ_n 的一个特征对：

$$\begin{aligned}\Sigma_n\widehat{f} &= \frac{1}{n}\sum_{i=1}^{n}(k_{X_i}\otimes k_{X_i})\left(\frac{1}{\widehat{\lambda}n}\sum_{j=1}^{n}k_{X_j}v_j\right)\\ &= \frac{1}{\widehat{\lambda}n}\sum_{i=1}^{n}\left(k_{X_i}\sum_{j=1}^{n}k(X_i,X_j)v_j\right) = \frac{1}{\widehat{\lambda}n}\sum_{i=1}^{n}k_{X_i}[K_n v]_i = \frac{1}{\widehat{\lambda}n}\sum_{i=1}^{n}k_{X_i}\widehat{\lambda}v_i = \widehat{\lambda}\widehat{f}.\end{aligned}$$

此外，f 是 k_{X_i} 的线性组合，因此属于 $\mathcal{H}$。 □

Σ_n 的总体版本是协方差算子

$$\Sigma = \mathbb{E}k_X \otimes k_X,$$

其中 $k_X = k(\cdot, X)$、$d = \mathbb{E}k_X$和 $X \sim \mathbb{P}$。稍后我们将证明，可以在适当的希尔伯特空间中定义此类随机元素的期望。在适当的假设下，可以证明，我们可以选择 $\{f_i\}_{i=1}^{K}$，即 Σ的前 K 个特征函数，使其取实值，并且在 $L^2(\mathcal{X},\mathbb{P})$中正交归一。然后，我们可以定义$V_1$：将 $\mathbb{C}^K \to \mathcal{H}$ 定义为 $V_1\alpha = \sum_{i=1}^{K}\alpha_i f_i$。类似地，我们可以根据（B.2）定义 $\widehat{V}_1$，其中 $\{\widehat{f}_i\}_{i=1}^{K}$是按照 K_n 中前 K 个正交归一特征向量的扩展。现在，我们可以应用一般理论证明以下结果，该结果类似于定理 3.2。

定理 B.2。在下文定义的一般假设下，存在由 $\mathcal{X}$、$\mathbb{P}$、$k(\cdot,\cdot)$确定的 C_6、C_7，使得当样本量 $n \geq C_6\tau$且对于某个 $\tau > 1$时，以 $1 - 6e^{-\tau}$的置信度，有

$$\inf\left\{\|V_1 - \widehat{V}_1 Q\|_{2\to\infty} : Q \in \mathbb{U}^K\right\} \leq C_7\frac{\sqrt{\tau}}{\sqrt{n}}. \tag{B.3}$$

定理 5.2 中提到的一般假设为

General Assumptions. 集合 $\mathcal{X}$ 是一个可分拓扑空间。核 k：$\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ 连续、对称、半正定，且

$$\sup_{x\in\mathcal{X}} k(x,x) < \infty. \tag{B.4}$$

将 Σ 视为从 $\mathcal{H}$ 到 $\mathcal{H}$ 的算子，其特征值满足 $\lambda_1 \geq \ldots \geq \lambda_K > \lambda_{K+1} \geq \ldots \geq 0$。$\Sigma_n$ 的前 K 个特征函数为 $\{f_i\}_{i=1}^{K} \subset C_b(\mathcal{X})$。

条件（B.4）确保我们所研究的所有算子都是 Hilbert-Schmidt 算子，并进一步保证有界随机元在 Hilbert-Schmidt 算子希尔伯特空间中的集中性。的可分性以及 $k(\cdot, \cdot)$ 的连续性保证了 RKHS $\mathcal{H}$ 是可分的，依据[34]的引理 4.33。

B.1. 证明概述

定理 B.2 的证明比定理 3.2 的证明更简单，因为我们可以直接使用与 $\mathcal{H}$ $k(\cdot,\cdot)$ 相关联的再生核希尔伯特空间。我们将证明，作为从 $\mathcal{H}$ 到 $\mathcal{H}$ 的算子，$\Sigma_n - \Sigma$ 的希尔伯特-施密特$\widehat{\ }$范数在 n 趋于无穷时趋于零。回顾一下，V_1 的列在 $L^2(\mathcal{X}, \mathbb{P}_n)$ 中仅是$\widetilde{\ }$正交归一的，但一般理论要求使用 V_1，其列在 $L^2(\mathcal{X}, \mathbb{P})$ 中正交归一。与证明定理 3.2 时面临的挑战类似，我们还需要处理在 $L^2(\mathcal{X}, \mathbb{P}_n)$ 和 $L^2(\mathcal{X}, \mathbb{P})$ 中正交归一性之间的区别所引入的误差，这是证明定理 3.2 的关键步骤之一。

严谨的处理将分为五个部分。第一部分，我们介绍 $\mathcal{L}_{HS}(\mathcal{H})$，即我们所使用的从 $\mathcal{H}$ 到 $\mathcal{H}$ 的希尔伯特-施密特算子所构成的希尔伯特空间，并论证 $\mathcal{H}$ 和 $\mathcal{L}_{HS}(\mathcal{H})$ 中随机元素的合理性。第二部分，我们在希尔伯特空间 $\mathcal{L}_{HS}(\mathcal{H})$ 中建立集中性结果。第三部分，我们检查一般理论所要求的其余条件。第四部分，我们将上述所有要素结合起来，并应用一般理论完成证明。

B.2. 第一部分：$\mathrm{L}_{\boldsymbol{HS}}(\mathrm{H})$ 空间以及 H 和 $\mathrm{L}_{\boldsymbol{HS}}(\mathrm{H})$ 中的随机元素

$\mathcal{L}_{HS}(\mathcal{H})$ 空间汇集了从 $\mathcal{H}$ 到 $\mathcal{H}$ 的所有希尔伯特-施密特算子，该空间本身也是一个希尔伯特空间。在说明协方差算子之前，我们首先$\widetilde{\ }$定义 $\mathcal{H}$ 中的均值元素以及 $\mathcal{L}_{HS}(\mathcal{H},\mathcal{H})$ 中的交叉协方差算子。

$\mathcal{H}$ 中的均值元素定义为 $\mu_X = \mathbb{E}_{X\sim\mathbb{P}}[k(\cdot, X)] \in \mathcal{H}$，使得$\langle f$，对于任意 $f \in \mathcal{H}$，都有 $\widetilde{\mu_k}\rangle_{\mathcal{H}} = \mathbb{E}_{X\sim\mathbb{P}}[f(X)]$。设 $(\mathcal{Y}, \mathcal{B}_\mathcal{Y}, \mathbb{Q})$ 是另一个概率$\widetilde{\ }$空间，且 $\mathcal{H}$ 是与核 $k(\cdot,\cdot)\widetilde{\ }$相关联的再生核希尔伯特空间，其中包含 $k(\cdot, y)$，$y \in \mathcal{Y}$ 作为其元素。交叉协方差算子 $C_{X,Y} =_\sim \mathbb{E}_{X\sim\mathbb{P},Y\sim\mathbb{Q}}[k(\cdot, X) \otimes k(\cdot, Y)]$ 是从 $\mathcal{H}$ 到 $\mathcal{H}$ 的希尔伯特 - 施密特算子，使得 $\widetilde{\ }$ 对于任意 $f \in \mathcal{H}$ 和 $g \in \mathcal{H}$，$\langle f, C_{X,Y} g\rangle_{\mathcal{H}} = \mathbb{E}_{X\sim\mathbb{P},Y\sim\mathbb{Q}}[f(X)g(Y)]$。最后，协方差算子为 $\Sigma = \mathbb{E}k_X \otimes k_X = C_{X,X} \in \mathcal{L}(\mathcal{H})$。

根据一般假设，该协方差算子确实是 Hilbert-Schmidt 算子，因为 $\|\mathbb{E}k_X \otimes k_X\|_{HS} \leq \mathbb{E}[\|k_X \otimes k_X\|^2_{HS}] = \mathbb{E}[k(X, X)^2] \leq (\sup_{x\in\mathcal{X}} k(x, x))^2 < \infty$。

B.3. 第二部分：希尔伯特空间 $\mathrm{L}_{HS}(\mathrm{H})$ 中的集中性

在第一部分中，我们已经证明 Σ 是从 $\mathcal{H}$ 到 $\mathcal{H}$的 Hilbert-Schmidt 算子。在本小节中，我们将利用引理 4.5证明 $\|\Sigma_n - \Sigma\|_{HS}$ 的集中性。

引理 B.3. 在一般假设下，以概率 $1 - 2e^{-\tau}$，有

$$\left\| \frac{1}{n} \sum_{i=1}^{n} k_{X_i} \otimes k_{X_i} - \mathbb{E} k_X \otimes k_X \right\|_{HS} \leq C_{23} \frac{\sqrt{\tau}}{\sqrt{n}} \tag{B.5}$$

其中某个常数 C_{23}。

证明。记 $M := \sqrt{\sup_{x \in \mathcal{X}} k(x,x)} < \infty$。对于任意 $i \in [n]$，有

$$\begin{aligned} \|k_{X_i} \otimes k_{X_i} - \mathbb{E} k_X \otimes k_X\|_{HS} &\leq \|k_{X_i} \otimes k_{X_i}\|_{HS} + \|\mathbb{E} k_X \otimes k_X\|_{HS} \\ &= \|k_{X_i}\|_{\mathcal{H}}^2 + \|\mathbb{E} k_X \otimes k_X\|_{HS} \\ &= k(X_i, X_i) + \|\mathbb{E} k_X \otimes k_X\|_{HS} \\ &\leq M^2 + \|\mathbb{E} k_X \otimes k_X\|_{HS}, \end{aligned}$$

这意味着 $k_{X_i} \otimes k_{X_i} - \mathbb{E} k_X \otimes k_X$ 是 $\mathcal{L}_{HS}(\mathcal{H}, \mathcal{H})$ 中有界的零均值独立随机变量。

我们可以取 $C_{23} = \sqrt{2}(M^2 + \|\mathbb{E} k_X \otimes k_X\|_{HS})$。然后根据引理 4.5，有

$$\left\| \frac{1}{n} \sum_{i=1}^{n} k_{X_i} \otimes k_{X_i} - \mathbb{E} k_X \otimes k_X \right\|_{HS} = \left\| \frac{1}{n} \sum_{i=1}^{n} (k_{X_i} \otimes k_{X_i} - \mathbb{E} k_X \otimes k_X) \right\|_{HS} \leq C_{23} \frac{\sqrt{\tau}}{\sqrt{n}}$$

概率为 $1 - 2e^{-\tau}$。 □

B.4. 第三部分：检查一般理论的条件

引理 B.4. 在一般条件下，以下事实成立：

1. 再生核希尔伯特空间 H 是 $L^2(\mathcal{X}, \mathbb{P})$的子空间。2. $\mathcal{H}$ 范数 $\|\cdot\|_{\mathcal{H}}$ 比无穷范数更强。3. Σ和 Σ_n 都是从 $\mathcal{H}$ 到 $\mathcal{H}$的希尔伯特-施密特算子。4. Σ 的所有特征值（计重数）都可以排列成一个非负实数的递减（可能无限）序列 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_K > \lambda_{K+1} \geq \ldots \geq 0$，且 λ_K 与 λ_{K+1}之间存在正间隔。5. 前 K 个特征函数 $\{f_i\}_{i=1}^{K} \subset \mathcal{H}$ 可以选取为 $L^2(\mathcal{X}, \mathbb{P})$中的函数正交归一集。6. Σ_n 有一个非递增的实非负特征值序列。

证明。回顾可知，一般条件意味着 $M := \sqrt{\sup_{x \in \mathcal{X}} k(x,x)} < \infty$。因此，核 $k(\cdot, \cdot)$ 是一个 Mercer 核，并满足条件 $\int_{\mathcal{X} \times \mathcal{X}} k^2(x,y) d\mathbb{P}(x) d\mathbb{P}(y) < \infty$。于是，第 1 项是 Mercer 定理的推论。

对于第 2 项，任取 $f \in \mathcal{H}$，根据再生性质和柯西-施瓦茨不等式，有 $|f(x)|^2 = \langle f, k(\cdot, x)\rangle_{\mathcal{H}}^2 \leq |k(x,x)|^2 \|f\|_{\mathcal{H}}^2 \leq M^2 \|f\|_{\mathcal{H}}^2$。因此 $\|f\|_\infty \leq M\|f\|_{\mathcal{H}}$。

引理证明的中间步骤中已检查第 3 项B.3。
第 4 项和第 5 项均由 Mercer 定理保证。

第 6 项成立，是因为 Σ_n 的谱与对称半正定核矩阵 K_n的谱之间的关系已在引理 B.1中得到验证。

□

B.5. 第四部分：整合所有要素

为了完成定理 B.2的证明，我们必须处理算子 $\widehat{V}_1$ 在 $L^2(\mathcal{X}, \mathbb{P}_n)$ 中只有正交归一列、而在 $L^2(\mathcal{X}, \mathbb{P})$ 中没有正交归一列所引入的误差。这可以使用与定理3.2证明第五部分中相同的技巧来完成。因此，此处省略技术性重复内容。