arXiv:2302.06757v1 [stat.ML] 13 Feb 2023

核化扩散映射

LOUCAS PILLAUD-VIVIEN[1,2] 和 FRANCIS BACH[3]

摘要。谱聚类 [NJW01] 和扩散映射 [CL06] 是广受认可的降维算法，基于与数据扩散结构相关的特征元素构建。此类方法的核心是通过图核方法 [HAL07], 近似拉普拉斯算子；然而，这种局部平均构造已知会受到高维诅咒的困扰 d。本文通过再生核希尔伯特空间方法构建一种不同的拉普拉斯算子估计量，该方法能够自然地适应问题的正则性。我们给出非渐近统计速率，证明所构建的核估计量能够规避维度诅咒。最后，我们讨论了可降低估计量计算成本且不损害其整体性能的技术（Nyström 子采样、傅里叶特征）。

1. 引言

基于再生核希尔伯特空间（RKHS）进行学习之所以成功，原因之一在于它能够自然地选取适合问题的测试函数基。更有意思的是，借助目标函数内在的正则性，RKHS 还能够规避维数灾难。这正是所有依赖局部平均的技术失效之处：逼近问题总会受到高维度的困扰 d，因为要取得良好效果，就需要 $n^{-1/d}$ 个点。这一主要差异也反映在核的性质上：逐点正核见于非参数估计文献 [Nad64]，而正半定（*PSD*）核则见于现代核学习 [SC08、SS02]。

本工作的核心，是使用 PSD 核来解决过去依赖局部方法处理的问题。具体而言，我们通过测度 μ 的主特征元素来估计相关的扩散算子（或拉普拉斯算子）。将其表述为一个无监督学习问题时，这可以视为一种基于数据扩散特性的降维技术。这正是著名的谱聚类算法 [VL07] 以及分子动力学领域中的扩散映射 [CL06] 的核心。然而，正如前文所述，这些算法基于图拉普拉斯算子；这是一种本质上局部的构造，在维度较高时扩展性较差 [HAL07]，也无法利用近期关于 PSD 核的研究成果，而这些研究正是针对潜在的高维场景展开的 [MT20、MCRR20]。

下面解释本文方法与图拉普拉斯算子方法之间的根本差异。当我们希望估计扩散算子（或其特征向量）时

$$\mathcal{L} := -\Delta + \langle \nabla V, \nabla \cdot \rangle, \tag{1}$$

困难之处之一在于逼近微分算子。尽管目前人们采用局部核平滑技术，但我们的方法有所不同。它利用了再生核希尔伯特空间中导数的再生性质以及 $\mathcal{L}$ 的自伴性来规避这一困难：这一策略在偏微分方程的数值分析中已取得了富有成效的结果，在那里被称为无网格方法 [SW06]。

从另一个角度来看，值得注意的是，[Sal98] 曾试图证明，考虑 $\mathcal{L}$ 的前几个特征向量是推广主成分分析方法 [Pea01，Hot33]的非线性方式是恰当的。当时，（i）无论是扩散和加权索伯列夫空间背后的理论，（ii）还是再生核希尔伯特空间理论，都尚未成熟。因此，他清楚地说明了（i）其分析的理论框架有限，但可以扩展；（ii）当时，由于必须对拉普拉斯算子进行离散化，因此在高维情况下不可能通过数值方法求解该问题。相当令人意外的是，关于图拉普拉斯算子的文献似乎忽略了 Salinelli 的开创性贡献。我们的工作可以被视为一种自然的

[1]纽约大学柯朗数学科学研究所，纽约，纽约州。[2]计算数学中心，弗莱铁研究所，西蒙斯基金会，纽约，纽约州。[3]法国国家信息与自动化研究所，巴黎高等师范学院，巴黎文理研究大学。

延续：我们借助现代工具进一步深化了对这种非线性主成分分析的理论理解，并提出了一种高效求解方法。

另请注意，本研究也受分子动力学应用的推动；在这一领域，扩散映射是一种重要的降维技术，可用于寻找反应坐标，即高维动力学中的慢扩散模态 [CBLK06]。本文也可视为对 [PVBL+20]的自然延伸；后者旨在估计 $\mathcal{L}$ 的第一个非零特征值（也就是说，其谱隙）。除研究更为成熟之外，本文的重点也截然不同：我们着重估计 $\mathcal{L}$ 的整个谱，并力求更精确地刻画其收敛性质。最后需要指出，我们将要介绍的过程可视为对由动力学生成的$\mathcal{L}$ [KNP+20]的 Koopman 生成元进行数据估计，并且充分利用了其自伴性。

2. 扩散算子

考虑概率测度 $d\mu$，定义在 $\mathbb{R}^d$ 上，其相对于勒贝格测度具有密度，并可写成如下形式：$d\mu(x) = e^{-V(x)}dx$，其中 V 称为势函数。考虑$H^1(\mu)$，即 $L^2(\mu)$中的函数子空间（也就是说，这些函数平方可积），且其所有一阶导数也属于 L^2，即 $H^1(\mu) = \{f \in L^2(\mu) \ \int_{\mathbb{R}^d} f^2 d\mu + \int_{\mathbb{R}^d} \|\nabla f\|^2 d\mu < \infty\}$，其中 ∇f 是 f 的梯度，$\|\cdot\|$ 是标准欧几里得范数。

本文旨在估计与测度 μ 相关联的扩散算子 $\mathcal{L}$，给定x_1、……、x_n根据 μ 独立同分布的样本。其定义为

$$\mathcal{L}\phi := -\Delta\phi + \nabla V \cdot \nabla\phi, \tag{2}$$

其中 ϕ 是一个足够光滑的测试函数。

2.1. 朗之万扩散。考虑 $\mathbb{R}^d$ 中的过阻尼朗之万扩散，即以下随机微分方程的解：

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \tag{3}$$

其中 $(B_t)_{t\geqslant 0}$ 是一个 d 维布朗运动。众所周知，[BGL14] $(X_t)_{t\geqslant 0}$ 的分布收敛于吉布斯测度 $d\mu$，并且庞加莱常数（见下文备注 2.1）控制着在 $L^2(\mu)$ 中趋于平衡的收敛速率。记由朗之万扩散 $(X_t)_{t\geqslant 0}$ 关联的马尔可夫半群为 $P_t(f)$。其定义如下：$P_t(f)(x) = \mathbb{E}[f(X_t)|X_0 = x]$。该半群满足以下动力学

$$\frac{d}{dt}P_t(f) = -\mathcal{L}P_t(f),$$

其中 $\mathcal{L}\phi = -\Delta\phi + \nabla V \cdot \nabla\phi$ 是一个微分算子，称为朗之万扩散的无穷小生成元（3）（Δ 表示 $\mathbb{R}^d$ 上的标准拉普拉斯算子）。注意，根据分部积分，半群$(P_t)_{t\geqslant 0}$关于 $d\mu$可逆，即：$\int f(\mathcal{L}g)\, d\mu = \int \nabla f \cdot \nabla g\, d\mu = \int (\mathcal{L}f) g\, d\mu$。这也表明 $\mathcal{L}$ 是作用于 $H^1(\mu)$上的对称正定算子。

注 2.1（与庞加莱常数的联系）。我们将 π 称为 $L^2(\mu)$ 在常函数上的正交投影算子：πf：$x \in \mathbb{R}^d \mapsto \int f d\mu$ 并定义 $L^2_0(\mu)$：$=$核 π。在假设 *0* 成立时（见下文），$\mathcal{L}$ 的第一个非零特征值为：

$$\mathcal{P}^{-1} = \inf_{f\in(H^1(\mu)\cap L^2_0(\mu))\setminus\{0\}} \frac{\langle f, \mathcal{L}f\rangle_{L^2(\mu)}}{\|f\|^2_{L^2(\mu)}}, \tag{4}$$

其中 $\mathcal{P}$ 也称为该分布的庞加莱常数 $d\mu$ [CSP+00]。

2.2. 扩散算子的一些有用性质。

半正定性。 我们看到的第一个性质是 $H^1(\mu)$ 中 $\mathcal{L}$ 的对称性和正定性。这源于以下分部积分恒等式：

$$\int f(\mathcal{L}g)\,d\mu = \int \nabla f \cdot \nabla g\,d\mu = \int (\mathcal{L}f)g\,d\mu, \tag{5}$$

这表明，由 $\mathcal{L}$ 诱导的二次型也是狄利克雷能量

$$\langle \mathcal{L}f, f\rangle_{L^2(\mu)} = \int \|\nabla f\|^2 d\mu =: \mathcal{E}(f).$$

与薛定谔算子的联系。在偏微分方程（PDEs）领域，如果一个算子是拉普拉斯算子与一个乘法算子之和，我们就称其为薛定谔型算子；之所以如此命名，是因为这类算子支配着量子系统的动力学 [HN05]。在这里，我们定义薛定谔算子 $\widetilde{\mathcal{L}} := -\Delta + \mathcal{V}$，其中 $\mathcal{V} := \frac{1}{2}\Delta V - \frac{1}{4}\|\nabla V\|^2$。我们可以证明，$\widetilde{\mathcal{L}}$ 和 $\mathcal{L}$ 彼此共轭：事实上，经过快速计算可得

$$\widetilde{\mathcal{L}} = e^{-V/2}\mathcal{L}\left[e^{V/2}\cdot\right].$$

由于薛定谔算子已得到充分研究，我们可以据此推断出关于$\mathcal{L}$的谱的一些有趣性质。事实上，

$$(\lambda, u) \text{ eigen-elements of } \widetilde{\mathcal{L}} \quad \Leftrightarrow \quad (\lambda, e^{V/2}u) \text{ eigen-elements of } \mathcal{L},$$

对于足够光滑的 f，我们还有如下等式：

$$\frac{\langle \mathcal{L}f, f\rangle_{L^2(\mu)}}{\|f\|^2_{L^2(\mu)}} = \frac{\langle \widetilde{\mathcal{L}}f, f\rangle_{L^2(\mathbb{R}^d)}}{\|f\|^2_{L^2(\mathbb{R}^d)}}.$$

谱 $\mathcal{L}$。由此我们可以推断出的最重要性质，是 $\mathcal{L}$的谱的性质。事实上，众所周知 [RS12]，若 $\mathcal{V}$ 局部可积、有下界且为强制的（$\mathcal{V}(x) \longrightarrow +\infty$，当 $\|x\| \to +\infty$时），则薛定谔算子具有紧预解式。特别地，本文始终采用以下假设

假设 0（$\mathcal{L}$的谱）。假设 $\frac{1}{2}\Delta V(x) - \frac{1}{4}\|\nabla V\|^2 \longrightarrow +\infty$，当 $\|x\| \to +\infty$.

假设0 意味着 $\mathcal{L}$ 具有紧预解式。这也意味着 $\mathcal{L}$ 具有纯离散谱，且其特征函数构成完备系。请注意，该假设还意味着扩散算子 $\mathcal{L}$ 存在谱隙，因而庞加莱不等式成立。在本文中，即使未明确说明，我们也始终采用假设0。关于 $\mathcal{L}$的谱，参见 [BGL14、HN05]。

3. Approximation of the diffusion operator in the RKHS

设 $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ 是一个以正定核 K 为核的再生核希尔伯特空间。假设如下：

假设 1 (普适性)。 $\mathcal{H}$ 在 $H^1(\mu)$中稠密。

请注意，对于大多数常见的核与分布组合，情况都是如此：高斯分布核、指数核具有普适性，[MXZ06] 当 μ 具有紧支撑或次高斯尾部时。由于式（2）中的扩散算子表达式涉及测试函数的导数，我们还需要再生核希尔伯特空间具备一些正则性。事实上，为了在我们的再生核希尔伯特空间中表示 ∇f，我们关键地利用了核空间的偏导数再生性质。为此，我们需要：

假设 2 (光滑性)。 K 是一个正定核，满足 $K \in \mathcal{C}^2(\mathbb{R}^d \times \mathbb{R}^d)$。

对于 $i \in [\![1,\ d]\!]$，记 $\partial_i = \partial_{x^i}$ 为关于 i的第 x个分量的偏导数算子。已有研究表明 [Zho08]，在假设2下，我们可以定义 $\mathcal{H} \ni \partial_i K_x :\ y \to \partial_{x^i} K(x,\ y)$，并且偏导数再生性质成立：$\forall f \in \mathcal{H}$且 $\forall x \in \mathbb{R}^d,\ \partial_i f(x) = \langle \partial_i K_x, f\rangle_{\mathcal{H}}$。因此，借助假设2，$\nabla f$ 很容易在再生核希尔伯特空间中表示。我们还需要核的一些有界性。

假设 *3*（有界性）。 K 是一个核，满足 $\forall x \in \mathbb{R}^d$, $K(x,x) \leqslant \mathcal{K}$ 且[1] $\|\nabla K_x\|^2 \leqslant \mathcal{K}_d$，其中 $\|\nabla K_x\|^2 := \sum_{i=1}^d \langle \partial_i K_x, \partial_i K_x \rangle = \sum_{i=1}^d \frac{\partial^2 K}{\partial x^i \partial y^i}(x,x)$（见下方计算）， x 和 y 分别表示 $(x,y) \mapsto K(x,y)$ 中的第一个和第二个变量。

表达式 $\|\nabla K_x\|^2$ 中的等式由以下计算得出：对所有 x， $y \in \mathbb{R}^d$, $\langle \partial_i K_x, \partial_i K_y \rangle = \partial_{x^i}(\partial_i K_y(x)) = \partial_{x^i}\partial_{y^i} K(x, y)$。注意，例如，高斯分布核和指数核满足假设1、2、3。此处陈述有界性是为了清晰起见；不过，除对数项外，若令 $\|K_X\|^2$, $\|\nabla K_X\|^2$ 为次高斯随机变量，本文的结果仍然成立。

例3.1（高斯分布核）。 一个典型例子是高斯核（或径向基函数），其带宽为 $\sigma > 0$；对于该核，我们可以计算，对于 $i \neq j$,

$$K(x,y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right), \qquad \partial_{x^i}\partial_{y^j} K(x,y) = -\frac{(x_i - y_i)(x_j - y_j)}{\sigma^4} K(x,y),$$
$$\partial_{x^i} K(x,y) = -\frac{(x_i - y_i)}{\sigma^2} K(x,y), \qquad \partial_{x^i}\partial_{y^i} K(x,y) = \left(\frac{1}{\sigma^2} - \frac{(x_i - y_i)^2}{\sigma^4}\right) K(x,y).$$

3.1. 将扩散算子嵌入再生核希尔伯特空间。 定义以下从 $\mathcal{H}$ 到 $\mathcal{H}$的算子：

$$\Sigma = \mathbb{E}_\mu\left[K_X \otimes K_X\right], \qquad \mathsf{L} = \mathbb{E}_\mu\left[\nabla K_X \otimes_d \nabla K_X\right], \tag{6}$$

其中， $\otimes$是标准张量积： $\forall f$， g， $h \in \mathcal{H}$, $(f \otimes g)(h) = \langle g, h \rangle_{\mathcal{H}} f$，而 $\otimes_d$定义如下：$\forall f$， $g \in \mathcal{H}^d$ 和 $h \in \mathcal{H}$, $(f \otimes_d g)(h) = \sum_{i=1}^d \langle g_i, h \rangle_{\mathcal{H}} f_i$。根据 K的再生性质，$(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$可通过算子 S 规范地嵌入 $(L^2(\mu), \langle \cdot, \cdot \rangle_{L^2(\mu)})$中，同时其伴随算子 S^* 定义为从 $L^2(\mu)$ 到 $\mathcal{H}$，满足对所有 $x \in \mathbb{R}^d$:

$$\forall f \in \mathcal{H}, \ \mathsf{S}f(x) = \langle f, K_x \rangle_{\mathcal{H}} = f(x), \qquad \forall f \in L^2(\mu), \ \mathsf{S}^* f(x) = \mathbb{E}_\mu\left[K(x,X) f(X)\right].$$

注意到 $\mathsf{S}^*\mathsf{S} = \Sigma$。根据这些定义，并利用式（ $\mathcal{L}$ （5）中推导出的对称性，我们可以在再生核希尔伯特空间中表示扩散算子 $\mathcal{L}$。

命题 3.2（ $\mathcal{L}$ 在再生核希尔伯特空间中的嵌入）。假设 *1*、*2* 成立，则

$$\mathsf{L} = S^* \mathcal{L} S, \tag{7}$$

其中，等号表示 $\mathcal{H}$中算子之间的相等。

证明。对于 $z \in \mathbb{R}^d$， $f \in \mathcal{H}$,

$$\begin{aligned}
\langle \mathsf{L}f, K_z \rangle_{\mathcal{H}} &= \int \nabla f(x) \cdot \nabla_x K(x,z) d\mu(x) \\
&= -\int \Delta f(x) K(x,z) d\mu(x) + \int \nabla f(x) \cdot \nabla V(x) K(x,z) d\mu(x) \\
&= \langle \mathcal{L}\mathsf{S}f, \mathsf{S}K_z \rangle_{L^2(\mu)} \\
&= \langle \mathsf{S}^* \mathcal{L} \mathsf{S} f, K_z \rangle_{\mathcal{H}},
\end{aligned}$$

因此算子之间相等。注意，从第一行到第二行时，我们用到了 $\mathcal{L}$的对称性。■

我们希望构造 $\mathcal{L}$ 在定义域 $H_0^1(\mu) := H^1(\mu) \cap L_0^2(\mu)$上的特征元素近似，其中回顾一下， $L_0^2(\mu) = \mathrm{Ker}\,\pi$ 表示不含常数的平方可积函数空间。同样，记 $\mathcal{H}_0 = (\mathrm{Ker}\mathsf{L})^\perp$为 $\mathcal{H}$ 中不含常值函数的子空间。注意，根据谱隙假设 0，该算子可逆。接下来，我们将近似 $\mathcal{L}^{-1}$的特征元素。首先，我们给出 $\mathcal{L}^{-1}$ 在再生核希尔伯特空间 $\mathcal{H}$中的表示，然后构造一个 $\mathcal{H}$ 上的算子，使其与 $\mathcal{L}^{-1}$具有相同的特征元素。事实上，若记L^{-1} 为L在$(\mathrm{Ker}\mathsf{L})^\perp$上的逆算子，则有：

[1]在 $\mathcal{K}_d$ 中，下标 d 表示该量预计会随 d （高斯核）线性缩放。

命题 3.3 ($\mathcal{L}^{-1}$的表示）。假设 *0*、*1*、*2* 成立，则

$$\mathcal{L}^{-1} = \mathsf{S}\mathsf{L}^{-1}\mathsf{S}^*, \tag{8}$$

*w*这里的等号表示定义域为 $H_0^1(\mu)$ 的算子之间的相等。

根据命题 3.3，我们可以将 $\mathcal{L}^{-1}$ 表示为再生核希尔伯特空间中的元素，通过嵌入 S实现。但我们真正希望的是在 $\mathcal{H}$ 上的一个算子，它与 $\mathcal{L}^{-1}$具有相同的特征元。这样的表示可以进行数值计算：这正是下面命题的目的。

定理 3.4 （ $\mathcal{L}^{-1}$ 作为再生核希尔伯特空间中函数的特征元）。将扩散算子的逆算子分解为 $\mathcal{L}^{-1} = \mathsf{S}\mathsf{L}^{-1}\mathsf{S}^* = \mathsf{S}\mathsf{L}^{-1/2}\mathsf{L}^{-1/2}\mathsf{S}^*$，则，

（*i*）$\mathsf{S}\mathsf{L}^{-1/2}$ 是从 $\mathcal{H}_0$ 到 $H_0^1(\mu)$ 的有界算子。（*ii*）$\mathsf{L}^{-1/2}\Sigma\mathsf{L}^{-1/2}$ 是 $\mathcal{H}_0$ 上的自伴紧算子，并且与 $\mathcal{L}^{-1}$具有相同的谱。（*iii*）如果 $\lambda \neq 0$ 是$\mathsf{L}^{-1/2}\Sigma\mathsf{L}^{-1/2}$ 的一个特征值，其特征向量为 $u \in \mathcal{H}_0$，那么 λ 是 $\mathcal{L}^{-1}$的一个特征值，其特征向量为$\mathsf{S}\mathsf{L}^{-1/2}u \in H_0^1(\mu)$。

该定理将使我们能够用有限样本集来近似 $\mathcal{L}^{-1}$ 的特征元，近似对象为算子 $\mathsf{L}^{-1/2}\Sigma\mathsf{L}^{-1/2}$(它仅在 $\mathcal{H}_0$ 上定义良好)。该定理的证明源于前一个命题中给出的 $\mathcal{L}^{-1}$ 的表示，以及一条关于希尔伯特算子的技术性引理，该引理证明见附录（引理 A.1）。

3.2. 估计量的定义。

经验算子。 我们定义 L 和 Σ 的经验对应量：将关于 μ 的期望替换为关于其经验测度 $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ 的期望，即可得到这些经验对应量，其中 $x_1, \ldots, x_n$ 是服从 $d\mu$分布的独立同分布样本。

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^n K_{x_i} \otimes K_{x_i}, \quad \text{and} \quad \widehat{\mathsf{L}} = \frac{1}{n}\sum_{i=1}^n \nabla K_{x_i} \otimes_d \nabla K_{x_i}. \tag{9}$$

因此，我们可能会倾向于将估计量定义为 $\widehat{\mathsf{L}}^{-1/2}\widehat{\Sigma}\widehat{\mathsf{L}}^{-1/2}$。然而，这一定义存在两个主要问题：

（i）如果 $f \in \mathrm{Ker}\,\widehat{\mathsf{L}}$，即对于所有 $i \leqslant n$，$\nabla f(X_i) = 0$，则 $\|\widehat{\mathsf{L}}^{-1/2}\widehat{\Sigma}\widehat{\mathsf{L}}^{-1/2}f\| = +\infty$。这是一个过拟合型问题。（ii）另一个问题与如下事实有关：求$\mathsf{L}^{-1/2}\Sigma\mathsf{L}^{-1/2}$ 的特征元素等价于求解广义特征值问题： $\Sigma f = \sigma\mathsf{L}f$。正如 [Cra76]所指出的，这类系统已知数值不稳定。用经验对应量替换算子时，这种情况尤其明显。这是一个稳定性问题。

正则化。这两个问题让人想起回归任务中过拟合的隐患。因此，与核岭回归类似，一个自然的想法是引入参数 λ进行正则化。由此得到以下估计量及其经验对应量的定义：

定义 3.5 (估计量的定义)。 在假设 *0*、*1*、*2*、*3*成立的条件下，我们定义逆扩散算子的两个估计量 $\mathcal{L}^{-1}$:

$$\textit{Biased estimator:} \qquad (\mathsf{L} + \lambda I)^{-1/2}\Sigma(\mathsf{L} + \lambda I)^{-1/2} \tag{10}$$

$$\textit{Empirical estimator:} \qquad (\widehat{\mathsf{L}} + \lambda I)^{-1/2}\widehat{\Sigma}(\widehat{\mathsf{L}} + \lambda I)^{-1/2}. \tag{11}$$

$\widehat{\ }$ $\widehat{\ }$下文为简化记号，定义 $\mathsf{L}_\lambda = \mathsf{L} + \lambda I$ 和 $\mathsf{L}_\lambda = \mathsf{L} + \lambda I$。显然，这种正则化的主要缺点是会给我们的估计带来偏差：更敏锐的读者会注意到， λ 越大，该问题就越接近核主成分分析 [MSS+99]。换句话说，λ 的尺度控制着我们希望从数据中提取的扩散信息的大小（这一观点还可进一步探讨，但我们暂且留待未来工作）。

在分析经验估计量的性能时，我们将特别关注它与标准算法的比较；该算法计算算子的特征元素：扩散映射 [CL06，HAL07]。我们强调，所提出的再生核希尔伯特空间方法能够充分利用再生核希尔伯特空间方法的诸多优点 [SS02]：统计方面，既能处理维数依赖问题，也能适应目标函数的正则性 [CDV07]，；计算方面，则可受益于文献中提出的技术，例如列子采样或使用随机特征 [MT20]。

3.3. 我们感兴趣的是近似哪些量?

问题的要求。 本研究自然且一般的目标是基于独立同分布样本，给出对扩散算子的近似。不过，实际上读者可能希望近似的对象还包括一些更具体的实用量：

- **整个算子**。 可以近似它在 $\mathcal{H}$ 中的表示， $H^1(\mu)$。正如第 2.2 小节所述，这可以用于估计薛定谔算子。还可以利用未标记数据的狄利克雷能量对半监督问题进行正则化，以利用其结构 [CPVBR21，CBB22]。
- **半群**。 事实上，由于 $\mathcal{L}$ 是动力学的无穷小生成元，我们也可以关注与之关联的半群的收敛性 $e^{t\mathcal{L}}$ [KNP+20]。
- **特征向量**。 由于这一估计量的主要应用之一，是通过 $\mathcal{L}$的特征向量计算数据的低维嵌入，因此我们直接关注特征向量的近似。既可以逐个近似特征向量，也可以近似由其中少数特征向量张成的有限维子空间。请注意，我们主要关注 $\mathcal{L}$的小特征值，它们对应于 $\mathcal{L}^{-1}$的大特征值，因为这些特征值决定了动力学的行为 [Lel13]。

- **特征值**。正如先前的研究已经处理了最大特征值 [PVBL+20]一样，我们希望近似一组特征值。另一项应用是构造用于聚类的扩散距离 [CL06]。

先前的结果。在先前的研究中，例如， [HAL07]和 [CL06] 证明了估计算子的收敛性。不过请注意，收敛定理给出的是逐点的结果，适用于有界域，且对维数的依赖关系很差，如 $n^{-1/d}$所示。我们将尝试克服这三项局限。请注意，向扩散算子的算子范数收敛意味着前文提到的所有量都收敛：**(i) 有限时间的半群**，这可由以下不等式得到： $\|e^B - e^A\| \leq \|B - A\| e^{\max\{\|A\|,\|B\|\}}$；**(ii) 特征向量和特征值，**这可直接通过扰动理论论证得出。值得注意的是，如果要近似 k维子空间，我们还将讨论更精细的界限，类似于 [ZB05]。

4. 估计量的统计分析

如前所述，为简化记号，对于算子 A，定义算子 $A_\lambda := A + \lambda I$。我们将问题分为两部分：偏差项和方差项

$$\left\| \widehat{\mathrm{L}}_\lambda^{-1/2} \widehat{\Sigma} \widehat{\mathrm{L}}_\lambda^{-1/2} - \mathrm{L}^{-1/2} \Sigma \mathrm{L}^{-1/2} \right\| \leqslant \underbrace{\left\| \widehat{\mathrm{L}}_\lambda^{-1/2} \widehat{\Sigma} \widehat{\mathrm{L}}_\lambda^{-1/2} - \mathrm{L}_\lambda^{-1/2} \Sigma \ \mathrm{L}_\lambda^{-1/2} \right\|}_{\text{方差}} + \underbrace{\left\| \mathrm{L}_\lambda^{-1/2} \Sigma \ \mathrm{L}_\lambda^{-1/2} - \mathrm{L}^{-1/2} \Sigma \mathrm{L}^{-1/2} \right\|}_{\text{bias}}$$

方差项对应于统计误差，其来源是我们只能获取分布 μ 的有限个 n 样本。偏差来自对算子 L 引入了由 λ 缩放的正则化。我们首先推导方差项的界。

4.1. 方差分析。

命题 4.1（统计误差分析）。 假设 *0*、*1*、*2*、*3*均成立。对于任意 $\delta \in (0, 1/3)$、$0 < \lambda \leqslant \|\mathsf{L}\|$ 以及任意整数 $n \geqslant 15 \frac{\mathcal{K}_d}{\lambda} \log^4 \mathrm{Tr}\mathsf{L}_\lambda \delta$，其概率至少为 $1 - 2\delta$，

$$\left\| \widehat{\mathsf{L}}_\lambda^{-1/2} \widehat{\Sigma} \widehat{\mathsf{L}}_\lambda^{-1/2} - \mathsf{L}_\lambda^{-1/2} \Sigma \mathsf{L}_\lambda^{-1/2} \right\| \leqslant \frac{8\mathcal{K}}{\lambda\sqrt{n}} \log(2/\delta) + \mathrm{o}\left(\frac{1}{\lambda\sqrt{n}} \right). \tag{12}$$

需要指出的是，命题 4.1 证明背后的分析并非全新：在 [PVBL+20]中，研究者使用类似工具考察了最大特征值的收敛性，主要区别在于，式（12）中的界是在算子范数下给出的。还请注意，为了表述清晰，我们只强调了 $\lambda\sqrt{n}$ 较大时的情形，而附录的引理 B.4 给出了一个明确的非渐近界。最后，我们强调：（i）该界无维数依赖；（ii）该界是**算子范数下的界**，对于算子收敛而言，这是一个强界，因为它蕴含许多其他结果：根据扰动理论的结论，可得特征值和特征向量收敛；还可得到相应半群的界，以及无限维算子的逐点收敛或其他形式的弱收敛。

4.2. **偏差分析**。 偏差分析较为困难，尽管现在所有对象都是确定性的。我们知道$\mathsf{L}_\lambda^{-1/2} \Sigma \mathsf{L}_\lambda^{-1/2}$ 是一个紧算子（Σ 是紧的，而$\mathsf{L}_\lambda^{-1/2}$ 有界），因此其谱是离散的，除 0外均由孤立点构成。同理，[RS12, 定理 XIII.67] 扩散算子的逆算子 $\mathcal{L}^{-1}$ 是紧的，因此我们可以讨论当 k 趋于 $\mathcal{L}^{-1}$ 时，$\mathsf{L}_\lambda^{-1/2} \Sigma \mathsf{L}_\lambda^{-1/2}$ 的第 λ 个特征元趋近于 0时的近似（如果特征值不是孤立的，则讨论特征子空间）。

估计量的一致性。 首先，如果我们只关心估计量的一致性，而不关心收敛速率，则有以下一致性结果：

命题 4.2（偏差收敛）。 在假设 *0*、*1*、*2*、*3*下，我们有以下算子范数意义下的收敛性：

$$\left\| \mathsf{L}_\lambda^{-1/2} \Sigma \mathsf{L}_\lambda^{-1/2} - \mathsf{L}^{-1/2} \Sigma \mathsf{L}^{-1/2} \right\| \underset{\lambda \to 0}{\longrightarrow} 0. \tag{13}$$

这一结果关键依赖于算子 $\mathsf{L}^{-1/2} \Sigma \mathsf{L}^{-1/2}$ 是紧算子这一事实（见定理3.4），并结合一些代数运算得出。

光滑特征函数的快速收敛率。若对分布 μ（以及再生核希尔伯特空间）没有更多先验知识，就很难推导出偏差关于正则化参数λ的普适收敛率。事实上，即使仅针对第一个特征值 $|\|\mathsf{L}_\lambda^{-1/2} \Sigma \mathsf{L}_\lambda^{-1/2}\| - \|\mathsf{L}^{-1/2} \Sigma \mathsf{L}^{-1/2}\||$推导定量扰动结果（它对应于该分布的庞加莱常数），也被认为是一个难题 [CSP+00]。这超出了本文的讨论范围。然而，在对分布作出标准假设的情况下，已知这些椭圆算子的特征函数是光滑的（通常要求吉布斯势光滑，且 μ 尾部 [BKRS22]快速衰减）。因此，与非参数回归中的做法类似，我们可以利用这一点，通过考察目标函数相对于再生核希尔伯特空间的光滑程度，来量化问题的难度 [CDV07] 。在这一领域，这通常称为源条件 [Die17]。在此，对于大多数应用 [CL06], ，我们希望逼近对应于 $\mathcal{L}^{-1}$ 最大特征值（ $\mathcal{L}$最小特征值）的 p-特征元，其中 $p \in \mathbb{N}^*$。下面对其光滑性作出一个自然的假设。

假设 4（问题的正则性）。 p 前几个特征向量 $\mathcal{L}$ 属于 $\mathcal{H}$.

一个典型例子是，对于任意 $p \in \mathbb{N}^*$，若考虑一个支撑集紧致的分布 Ω，其密度为 $\mu = e^{-V} \in \mathcal{C}^\infty(\Omega)$，并使用高斯核，就会出现这种情况。

我们记作 Π^p： $\mathcal{H} \to \mathcal{H}$ 投影到 p L的最大特征向量所张成空间上的谱投影算子$^{-1/2} \Sigma \mathsf{L}^{-1/2}$。严格来说，假设 4 意味着对所有 $v \in$ 张成空间 Π^p,

$$\|\mathsf{L}^{-1/2} v\|_{\mathcal{H}} < +\infty,$$

或者等价地用算子表示： $\|\Pi^p \mathsf{L}^{-1/2} \Pi^p\| < \infty$。事实上，得益于定理3.4-(iii)，对于这样的 $v \in \mathcal{H}$，$S\mathsf{L}^{-1/2}v \in \mathcal{H}$，因此根据等距性，这意味着 $\|(SS^*)^{-1/2} S\mathsf{L}^{-1/2}v\|_{L^2} < +\infty$，这等价于上述条件。

命题 4.3 （源条件下的快速收敛率）。 在假设*0*、*1*、*2*、*3*、*4*下，我们有如下的算子范数界：

$$\left\| \Pi^p \left(\mathsf{L}_\lambda^{-1/2} \Sigma \mathsf{L}_\lambda^{-1/2} - \mathsf{L}^{-1/2} \Sigma \mathsf{L}^{-1/2} \right) \Pi^p \right\| \leq 2\lambda \mathcal{P} \|\Pi^p \mathsf{L}^{-1/2} \Pi^p\|. \tag{14}$$

该命题意味着，在源条件下，前 p 个特征向量的偏差与 λ 呈线性关系。此外，界中还出现了两个关键的正则性假设，这一点非常值得注意：（i）通过其庞加莱常数 $\mathcal{P}$ 刻画的测度 μ 的复杂性度量；（ii）关于我们希望近似的目标特征向量的光滑 先验 。为清晰起见，这里我们展示了目标属于再生核希尔伯特空间（RKHS）的情形，但需要指出的是，在更精确（且更具技术性）的源假设下，可以很容易地根据这一结果调整出更精细的界。

4.3. 源假设下的一致性与收敛率。 为总结前两节的结果与讨论，我们在此陈述估计量的整体一致性，以及关于数据的经验估计量的最终界。

定理 4.4 （一致性与收敛率）。 在假设 *0*、*1*、*2*、*3*成立的条件下，对于任意 $\delta \in (0, 1/2)$ 以及任意整数 $n \geqslant \mathsf{K} \log \frac{1}{\delta}$，其中 K 依赖于 $\mathcal{K}$，$\mathcal{K}_d$ 以至少 $1-2\delta$ 的概率，取一列满足 $\lambda_n \to 0$ 且 $\lambda_n \sqrt{n} \to +\infty$ 的正则化项，则我们的估计量是一致的

$$\widehat{\mathsf{L}}_{\lambda_n}^{-1/2} \widehat{\Sigma} \widehat{\mathsf{L}}_{\lambda_n}^{-1/2} \underset{n\to\infty}{\longrightarrow} \mathsf{L}^{-1/2} \Sigma \mathsf{L}^{-1/2}, \tag{15}$$

其中收敛在算子范数意义下成立。此外，假设 *4*成立，则若 $\lambda = \mathcal{K} n^{-1/4}$，

$$\left\| \Pi^p \left(\widehat{\mathsf{L}}_{\lambda_n}^{-1/2} \widehat{\Sigma} \widehat{\mathsf{L}}_{\lambda_n}^{-1/2} - \mathsf{L}^{-1/2} \Sigma \mathsf{L}^{-1/2} \right) \Pi^p \right\| \leq \frac{8 + 2\mathcal{K}\,\mathcal{P} \|\Pi^p \mathsf{L}^{-1/2} \Pi^p\|}{n^{1/4}} + \mathrm{o}\left(n^{-1/4} \right), \tag{16}$$

其中， $\forall p \in \mathbb{N}^*$, Π^p 是由 $\mathsf{L}^{-1/2} \Sigma \mathsf{L}^{-1/2}$ 的最大 p 个特征向量张成的谱投影算子。

该定理量化了所构造估计量的统计性能：我们强调，在光滑性假设 4下，收敛到 $\mathcal{L}$ 的任意估计特征函数的收敛率都不依赖于维数。这与 $n^{-1/d}$ 图拉普拉斯算子/扩散映射的收敛率形成对比。这种差异反映了局部逼近技术与核方法之间更普遍且内在的区别：它们会根据问题的正则性进行自适应。

5. 估计量的数值构造

除了统计性能之外，核方法还具有良好的数值策略，可以在保持总体精度的同时降低计算成本 [MT20，第 19]节。我们非正式地讨论如何在我们的背景下应用这些方法。

计算估计量：朴素方法。为了计算估计量 $\widehat{\mathsf{L}}_\lambda^{-1/2} \widehat{\Sigma} \widehat{\mathsf{L}}_\lambda^{-1/2}$ ，需要能够表示这些算子

$$\widehat{\Sigma} = \frac{1}{n} \sum_{i=1}^n K_{x_i} \otimes K_{x_i}, \qquad \widehat{\mathsf{L}} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^d \partial_j K_{x_i} \otimes \partial_j K_{x_i},$$

为清晰起见，此处重述其表达式。事实上，只需在张成空间 $\{K_{x_i}\}_{i \leq n}$+Span$\{\partial_j K_{x_i}\}_{i \leq n,\ j \leq d}$上表示它们。一旦构造出这样的矩阵 $(\Sigma,\ L) \in \mathbb{R}^{(n+nd)\times(n+nd)s}$，就可以通过求解**广义特征值问题**来高效地计算该算子：即求出所有 $(\psi_k,\ \mu_k) \in \mathbb{R}^{n+nd} \times \mathbb{R}$，对于 $k \in [\![1,\ n+nd]\!]$，使得

$$\Sigma \psi_k = \mu_k L_\lambda \psi_k, \tag{17}$$

然后，借助特征值分解 $\widehat{\mathrm{L}}_\lambda^{-1/2}\widehat{\Sigma}\widehat{\mathrm{L}}_\lambda^{-1/2} = \sum_{k=1}^{n+nd}\mu_k f_k \otimes f_k$，我们可以写成

$$f_k = \sum_{i=1}^{n}\psi_k[i]K_{x_i} + \sum_{i=1}^{n}\sum_{j=1}^{d}\psi_k[n+ij]\partial_j K_{x_i}.$$

显然，瓶颈之一在于构造大型矩阵 Σ 和 L，它们分别约有 n^2d^2 个系数，然后求解与之相关的广义特征值问题。因此，当 n 或 d 过大时，该过程将变得难以处理。幸运的是，我们有一种可行的解决方案：要实现核方法，我们只需要对核矩阵进行近似。下面我们提出两种成熟的方法，用于减少计算量。

Nyström 近似/列子采样。该方法的思路是通过仅选择其中的Σ, L 列来构造它们的低秩近似。请注意，在有利的情况下，$p \in \mathbb{N}^*$ 可以低至 $p\log(n)$，而不会损害 [统计性能15]RCR。我们仅选择 p 列，这些列对应于元素$(K_{x_i})_{i\leq p}$。下面的算法出于其他目的发表于 [CPVBR21], ，它返回我们希望近似的特征向量：在算法1结束时，对于 $k \leq p$，第 k个近似特征值为 μ_k，并且

算法 1： 使用 Nyström 方法计算特征向量

数据： $(x_i)_{i\leq n}$、一个核 k 和一个正则化项 λ计算 $S_p = (k(x_i,x_l))_{i\leq n,l\leq p} \in \mathbb{R}^{n\times p}$；计算 $D_p = (\partial_{1,j}k(x_i,x_l))_{(i\leq n,j\leq d),l\leq p} \in \mathbb{R}^{nd\times p}$；构造 $\Sigma_p = S_p^\top S_p \in \mathbb{R}^{p\times p}$ 和 $\mathrm{L}_p = D_p^\top D_p \in \mathbb{R}^{p\times p}$；求得 $(\psi_k,\mu_k)_{k\leq p}$，即 $(\Sigma_p,\mathrm{L}_p+\lambda I_p)$ 的广义特征元素；

其对应的特征函数写作 $f_k(x) = \sum_{i=1}^{p}\Psi_k[i]K(x_i,x)$，其中 $\Psi_k = (\mathrm{L}_p+\lambda I_p)^{-1/2}\psi_k$。从数值复杂度来看，主要开销来自于在 $\mathcal{O}(p^2nd)$ 中构造 L^p，以及在 $\mathcal{O}(p^3)$ 中求广义特征元素。在算法过程中，我们使用记号 $\partial_{1,j}k(x_i,x_l)$，以强调导数应作用于第一个变量。

随机特征。随机特征 [RR08] 是另一种规避这一问题的方法：通过显式构造特征来近似任意平移不变核 $k(x,y) = k(x-y)$。更准确地说，设 $p \in \mathbb{N}^*$ 为随机特征的数量，$(w_l)^{l\leq p}$ 为一组随机变量，它们相互独立且同分布，服从 $\mathbb{P}(dw) = \int_{\mathbb{R}^d} \mathrm{e}^{-\mathrm{i}w^\top\delta}K(\delta)d\delta\,dw$，而 $(b_l)^{l\leq p}$ 相互独立且服从 $[0,2\pi]$ 上的均匀分布，则特征向量 $\phi_p(x) = \sqrt{\frac{2}{p}}\left(\cos(w_1^\top x+b_1)\ldots,\cos(w_M^\top x+b_M)\right)^\top \in \mathbb{R}^M$ 满足 $K(x,x') \approx \langle\phi_p(x)\ \phi_p(x')\rangle_2$。因此，随机特征可以用 $p\times p$ 矩阵来近似 Σ 和 $\bar{\mathrm{L}}$。在算法 2 的末尾，对于 $k \leq p$，第 k 个近似特征值为 μ_k，其对应的

算法 2： 使用随机特征计算特征向量

数据： $(x_i)_{i\leq n}$ 和一个正则化项 λ计算 $S_p = (\cos(w_l^\top x_i+b_l))_{i\leq n,\ l\leq p} \in \mathbb{R}^{n\times p}$；计算 $D_p^j = (-w_l[j]\sin(w_l^\top x_i+b_l))_{i\leq n,\ l\leq p} \in \mathbb{R}^{n\times p}$，对所有 $j\leq d$；构建 $D_p = \sum_{j=1}^{d}D_p^j \in \mathbb{R}^{n\times p}$；构建 $\Sigma_p = S_p^\top S_p \in \mathbb{R}^{p\times p}$ 和 $\mathrm{L}_p = D_p^\top D_p \in \mathbb{R}^{p\times p}$；求出 $(\psi_k,\ \mu_k)_{k\leq p}$的广义特征元素（$\Sigma_p$，$\mathrm{L}_p+\lambda I_p$）；

特征函数写作 $f_k(x) = \langle\Psi_k,\phi_p(x)\rangle$，其中 $\Psi_k = (\mathrm{L}_p+\lambda I_p)^{-1/2}\psi_k$。与之前类似，主要开销来自在 $\mathcal{O}(p^2nd)$ 中构建 L^p，以及在 $\mathcal{O}(p^3)$ 中求广义特征元素。

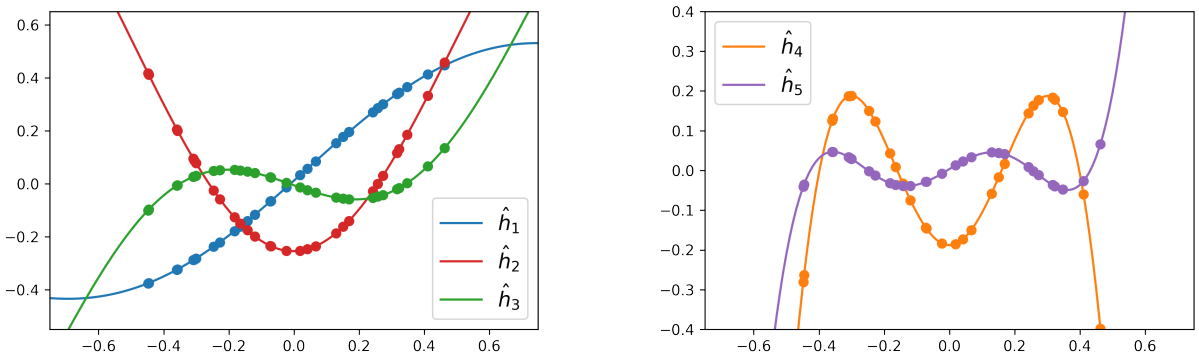

图 1. Ornstein-Uhlenbeck 算子的前五个特征函数的估计：厄米多项式。

厄米多项式。 作为本节数值分析的总结，并为展示结果，我们给出一个典型示例，其中 $\mathcal{L}$ 的特征函数是已知的；我们对其进行估计。具体而言，取 $\mu(x) = e^{-x^2/2}$，即一维高斯分布。于是 $\mathcal{L}f = f'' - xf'$ 是 奥恩斯坦–乌伦贝克 算子，其特征函数已知为厄米多项式 [BGL14]。借助算法 1，我们用该方法估计前五个厄米多项式，其中 $n = p = 30$，正则化参数为 $\lambda = 0.1$ 且使用高斯核。结果见图 1。数据点以圆点表示，构建出的特征函数以实线绘出。请注意，该近似仅在 $H^1(e^{-x^2/2})$上有效，因此估计的ˆ特征函数在数据集之外表现不佳：这一点的显著例子是 h_1的表现，它在数据集区间上是线性函数，但在没有数据的地方会迅速偏离。

6. 结论与进一步思考

与图拉普拉斯算子的比较。在这项工作中，我们证明了可以估计扩散算子的特征元。该构造依赖于半正定核方法，而此前的方法依赖于局部平均技术。与图拉普拉斯算子相比，这使得在高维情况下能够进行高效估计 [HAL07]。更确切地说，在目标特征向量具有光滑性的条件下，我们证明了再生核希尔伯特空间估计量的统计收敛速率不依赖于维度。最后，我们讨论了基于 Nyström 近似 [WS00] 或傅里叶特征 [RR08]构造这些特征元的高效计算方法。不过，本文侧重于估计量的数学基础及其统计性能；对当前使用的图拉普拉斯算子、扩散映射和谱聚类进行精确的计算与实验比较，将留待未来工作。

核的选择。 我们只作了简要讨论的另一个问题是核的选择。事实上，在涉及具体应用时，核工程是一项核心问题。对于这个问题，正如源条件所强调的，应选择再生核希尔伯特空间，使其能够很好地逼近 $H^1(\mu)$ 中扩散算子的特征函数。一般而言， $H^1(\mu)$ 与 $\mathcal{H}$ 之间的相互作用是再生核希尔伯特空间逼近理论的核心所在，也是一个根本性问题；理解这种联系，应能在大多数情况下实现无维度的方法。

马尔可夫链。应用概率领域近期有一系列研究，旨在根据马尔可夫链的前 n 次迭代 [HKS15]来估计其谱隙。估计量的第一个特征值似乎恰好也能实现同样的功能，值得进一步探究我们的算法与他们的方法之间的区别。这需要对我们的算法稍作调整，并将样本独立同分布的假设改为马尔可夫假设。

参考文献

[BBG20] Raphaël Berthier、Francis Bach 和 Pierre Gaillard。无噪声线性模型下随机梯度下降的紧非参数收敛率，2020。[BGL14] Dominique Bakry、Ivan Gentil 和 Michel Ledoux。《马尔可夫扩散算子的分析与几何》。施普林格，2014。[BKRS22] Vladimir I. Bogachev、Nicolai V. Krylov、Michael Röckner 和 Stanislav V. Shaposhnikov。《福克-普朗克-科尔莫戈罗夫方程》，第 207 卷。美国数学会，2022。[CBB22] Vivien Cabannes、Alberto Bietti 和 Randall Balestriero。论无监督表征学习中的最小变化。arXiv 预印本 arXiv:2211.03782，2022。[CBLK06] Ronald Coifman、Nadler Boaz、Stéphane Lafon 和 Ioannis Kevrekidis。动力系统的扩散映射、谱聚类与反应坐标。《应用与计算调和分析》，21(12):113–127，2006。[CDV07] Andrea Caponnetto 和 Ernesto De Vito。正则化最小二乘算法的最优收敛率。《计算数学基础》，7(3):331–368，2007。[CL06] Ronald R. Coifman 和 Stéphane Lafon。扩散映射。《应用与计算调和分析》，21(1)，2006。[CPVBR21] Vivien Cabannes、Loucas Pillaud-Vivien、Francis Bach 和 Alessandro Rudi。通过半监督学习中的拉普拉斯正则化克服维数灾难。《神经信息处理系统进展》，34:30439–30451，2021。[Cra76] Charles R Crawford。一个稳定的广义特征值问题。《SIAM 数值分析杂志》，13 860，1976 6 854( ): – 。[CSP+00] Ané Cécile、Blachère Sébastien、Fougères Pierre、Gentil Ivan、Malrieu Florent、Roberto Cyril 和 Scheffer Grégory。《苏博列夫对数不等式》。全景与综述 10，法国数学会（SMF），2000。[Die17] Aymeric Dieuleuveut。《希尔伯特空间中的随机逼近》。博士论文，巴黎高等师范学院 - 法国国家信息与自动化研究所，2017。[HAL07] Matthias Hein、Jean-Yves Audibert 和 Ulrike von Luxburg。图拉普拉斯算子及其在随机邻域图上的收敛性。《机器学习研究杂志》，第 1325–1368 页，2007。[HKS15] Daniel Hsu、Aryeh Kontorovich 和 Csaba Szepesvári。基于单条样本路径的可逆马尔可夫链混合时间估计。载于《神经信息处理系统进展》，第 1459–1467 页，2015。[HN05] Bernard Helffer 和 Francis Nier。《福克-普朗克算子与维滕拉普拉斯算子的次椭圆估计和谱理论》。《数学讲义》，1862，2005。[Hot33] Harold Hotelling。将一组统计变量分解为主成分。《教育心理学杂志》，24(6):417，1933。[KNP+20] Stefan Klus、Feliks Nüske、Sebastian Peitz、Jan-Hendrik Niemann、Cecilia Clementi 和 Christof Schütte。数据驱动的 Koopman 生成元近似：模型降阶、系统辨识与控制。《Physica D：非线性现象》，406:132416，2020。[Lel13] Tony Lelièvre。分析亚稳随机过程的两种数学工具。载于《2011 年数值数学与高级应用》，第 791–810 页，柏林、海德堡，2013。施普林格。[MCRR20] Giacomo Meanti、Luigi Carratino、Lorenzo Rosasco 和 Alessandro Rudi。核方法的规模再创新高：高效处理数十亿个点。载于《神经信息处理系统进展》，2020。[MSS+99] Sebastian Mika、Bernhard Schölkopf、Alex J. Smola、Klaus-Robert Müller、Matthias Scholz 和 Gunnar Rätsch。特征空间中的核主成分分析与去噪。载于《神经信息处理系统进展》，第 536–542 页，1999。[MT20] Per-Gunnar Martinsson 和 Joel A. Tropp。随机化数值线性代数：基础与算法。《数值学报》，29:403–572，2020。[MXZ06] Charles A Micchelli、Yuesheng Xu 和 Haizhang Zhang。通用核。《机器学习研究杂志》，7:2651–2667，2006。[Nad64] Elizbar A. Nadaraya。论回归估计。《概率论及其应用》，9(1):141–142，1964。[NJW01] Andrew Ng、Michael Jordan 和 Yair Weiss。论谱聚类：分析与算法。载于《神经信息处理系统进展》，14，2001。[Pea01] Karl Pearson。论空间中点系的最佳拟合直线与平面。《伦敦、爱丁堡和都柏林哲学杂志与科学杂志》，2(11):559–572，1901。[PV20] Loucas Pillaud-Vivien。《再生核希尔伯特空间中的学习：随机梯度下降与拉普拉斯估计》。博士论文，巴黎文理研究大学，2020。[PVBL+20] Loucas Pillaud-Vivien、Francis Bach、Tony Lelièvre、Alessandro Rudi 和 Gabriel Stoltz。庞加莱常数的统计估计及其在多模态分布采样中的应用。载于《人工智能与统计国际会议》，第 2753–2763 页，2020。[PVRB18] Loucas Pillaud-Vivien、Alessandro Rudi 和 Francis Bach。通过多次遍历解决困难学习问题时随机梯度下降的统计最优性。《神经信息处理系统进展》，31，2018。[RCR15] Alessandro Rudi、Raffaello Camoriano 和 Lorenzo Rosasco。少即是多：Nyström 计算正则化。载于《神经信息处理系统进展》，第 1657–1665 页，2015。[RR08] Ali Rahimi 和 Benjamin Recht。适用于大规模核机器的随机特征。载于《神经信息处理系统进展》，第 1177–1184 页，2008。

[RR17] Alessandro Rudi 和 Lorenzo Rosasco。随机特征学习的泛化性质。载于《神经信息处理系统进展》，第 3218–3228 页，2017 年。[RS12] Michael Reed 和 Barry Simon。《现代数理物理方法：泛函分析》，第 IV 卷。Elsevier，2012 年。[Sal98] Ernesto Salinelli。非线性主成分 I：具有正有界密度的绝对连续随机变量。《统计年刊》，26(2)：596–616，1998 年。[SC08] Ingo Steinwart 和 Andreas Christmann。《支持向量机》。施普林格科学与商业媒体，2008 年。[SS02] B. Schölkopf 和 A. J. Smola。《核方法学习》。麻省理工学院出版社，2002 年。[Sta17] Minsker Stanislav。自伴算子的伯恩斯坦不等式若干扩展。《统计与概率快报》，127：111–119，2017 年。[SW06] Robert Schaback 和 Holger Wendland。核技术：从机器学习到无网格方法。《数值学报》，15：543，2006 年。[Tro12] Joel A. Tropp。随机矩阵的易用工具：简介。NIPS 教程，2012 年。[VL07] Ulrike Von Luxburg。谱聚类教程。《统计与计算》，17(4)：395–416，2007 年。[VPVF21] Aditya Vardhan Varre、Loucas Pillaud-Vivien 和 Nicolas Flammarion。插值区间内最小二乘随机梯度下降的末次迭代收敛。《神经信息处理系统进展》，34：21581–21591，2021 年。[WS00] Christopher Williams 和 Matthias Seeger。利用 Nyström 方法加速核机器。《神经信息处理系统进展》，13，2000 年。[Yur95] Vadim Vladimirovich Yurinsky。和的分布的高斯及相关近似，第 163–216 页。施普林格柏林海德堡，1995 年。[ZB05] Laurent Zwald 和 Gilles Blanchard。核主成分分析中特征子空间的收敛性。《神经信息处理系统进展》，18，2005 年。[Zho08] Ding-Xuan Zhou。学习理论中核方法的导数再生性质。《计算与应用数学杂志》，220(1)：456–463，2008 年。

附录

请注意，这就是全部附录。为避免符号繁琐，除非确有必要，我们不会区分 $\mathcal{H}_0$ 与 $\mathcal{H}$以及 $H_0^1(\mu)$ 和 $H^1(\mu)$。

附录 A. 关于扩散算子及其逆的嵌入的证明 E

A.1. **命题 3.3 的证明。** 我们要证明，在 $H_0^1(\mu)$ 上， $\mathcal{L}^{-1} = \mathsf{S}\mathsf{L}^{-1}\mathsf{S}^*$。

命题的证明 命题 *3.3*。 令 $g \in \operatorname{Ran}\mathsf{S}$，存在 $f \in \mathcal{H}$ 使得 $g = \mathsf{S}f$。我们来计算：

$$\mathsf{S}\mathsf{L}^{-1}\mathsf{S}^*\mathcal{L}g = \mathsf{S}\mathsf{L}^{-1}\mathsf{S}^*\mathcal{L}\mathsf{S}f = \mathsf{S}\mathsf{L}^{-1}\mathsf{L}f = \mathsf{S}f = g.$$

此外，由于 $\mathcal{L}$在 $\operatorname{Ran}\mathsf{S} \cap H^1(\mu) \cap L_0^2(\mu)$ 上可逆，左逆与右逆相同。因此， $\mathcal{L}^{-1}$和 $\mathsf{S}\mathsf{L}^{-1}\mathsf{S}^*$ 在 $\operatorname{Ran}\mathsf{S}$ 上相等。此外，我们可以注意到， $\mathcal{L}^{-1}$ 和 $\mathsf{S}\mathsf{L}^{-1}\mathsf{S}^*$ 在 $L^2(\mu)$ 上有界。事实上，

$$\begin{aligned}\mathcal{P} = \sup_{f\in(\operatorname{Ker}\mathsf{L})^\perp} \frac{\langle f, \mathsf{S}^*\mathsf{S}f\rangle_{\mathcal{H}}}{\langle f, \mathsf{L}f\rangle_{\mathcal{H}}} \geqslant \sup_{f\in(\operatorname{Ker}\mathsf{L})^\perp} \frac{\langle \mathsf{L}^{-1/2}f, \mathsf{S}^*\mathsf{S}\mathsf{L}^{-1/2}f\rangle_{\mathcal{H}}}{\langle \mathsf{L}^{-1/2}f, \mathsf{L}\mathsf{L}^{-1/2}f\rangle_{\mathcal{H}}} &= \sup_{f\in(\operatorname{Ker}\mathsf{L})^\perp} \frac{\langle f, \mathsf{L}^{-1/2}\mathsf{S}^*\mathsf{S}\mathsf{L}^{-1/2}f\rangle_{\mathcal{H}}}{\|f\|_{\mathcal{H}}^2} \\ &= \|\mathsf{L}^{-1/2}\mathsf{S}^*\mathsf{S}\mathsf{L}^{-1/2}\|_{\mathcal{H}} \\ &= \|\mathsf{S}\mathsf{L}^{-1}\mathsf{S}^*\|_{L^2(\mu)}.\end{aligned}$$

As $\mathcal{L}^{-1}$ 和 $\mathsf{S}\mathsf{L}^{-1}\mathsf{S}^*$ 在 $\operatorname{Ran}\mathsf{S}$ 上相等且连续，因此在其闭包上也相等。 ■

A.2. **定理 3.4 的证明：借助希尔伯特空间之间算子的一个技术性结果。** 下面的引理通过考虑 $A = S\mathsf{L}^{-1/2}$, $\mathcal{H}_1 = \mathcal{H}_0$ 和 $\mathcal{H}_2 = H_0^1(\mu)$，给出定理 3.4 的证明。

引理 A.1 (A*A 与 AA* 之间的联系：紧情形。)。 设 $\mathcal{H}_1$ 和 $\mathcal{H}_2$ 为两个希尔伯特空间。设 A 是一个从 $\mathcal{H}_1$ 到 $\mathcal{H}_2$ 的算子，使得 A*A 是 $\mathcal{H}_1$上的自伴紧算子。那么，

(i) A 是一个从 $\mathcal{H}_1$ 到 $\mathcal{H}_2$的有界算子。*(ii)* AA* 是 $\mathcal{H}_2$ 上的自伴紧算子，且与AA*具有相同的谱。*(iii)* 若 $\lambda \neq 0$ 是A*A 的一个特征值，对应的特征向量为 $u \in \mathcal{H}_1$，则 λ 是AA* 的一个特征值，对应的特征向量为$\mathsf{A}u \in \mathcal{H}_2$。

证明。 首先注意到，A 必然是有界的。事实上，令 $u \in \mathcal{H}_1$,

$$\|\mathsf{A}u\|_{\mathcal{H}_2}^2 = \langle \mathsf{A}u, \mathsf{A}u\rangle_{\mathcal{H}_2} = \langle \mathsf{A}^*\mathsf{A}u, u\rangle_{\mathcal{H}_1} \leqslant \|\mathsf{A}^*\mathsf{A}u\|_{\mathcal{H}_1}\|u\|_{\mathcal{H}_1} \leqslant \|\mathsf{A}^*\mathsf{A}\|\,\|u\|_{\mathcal{H}_1}^2.$$

因此， $\|\mathsf{A}\| \leqslant \sqrt{\|\mathsf{A}^*\mathsf{A}\|}$。

其次，由于 A*A 在 $\mathcal{H}_1$上是自伴紧算子，存在 $(\psi_i)_{i\in\mathbb{N}}$ 标准正交基 $\mathcal{H}_1$ 以及实数序列$(\lambda_i)_{i\in\mathbb{N}}$，使得：

$$\mathsf{A}^*\mathsf{A} = \sum_{i\geqslant 0} \lambda_i \psi_i \otimes \psi_i,$$

其中，无穷级数表示算子的强收敛。现在，在左侧复合 A* 、在右侧复合 A*，得到：

$$(\mathsf{A}\mathsf{A}^*)^2 = \mathsf{A}\mathsf{A}^*\mathsf{A}\mathsf{A}^* = \sum_{i\geqslant 0} \lambda_i (\mathsf{A}\psi_i)\otimes(\mathsf{A}\psi_i) = \sum_{i\geqslant 0} \lambda_i^2 (\mathsf{A}\frac{\psi_i}{\sqrt{\lambda_i}})\otimes(\mathsf{A}\frac{\psi_i}{\sqrt{\lambda_i}}).$$

因此， $\mathsf{A}\mathsf{A}^* = \sum_{i\geqslant 0} \lambda_i (\mathsf{A}\frac{\psi_i}{\sqrt{\lambda_i}})\otimes(\mathsf{A}\frac{\psi_i}{\sqrt{\lambda_i}})$, 且是一个紧算子。当然，我们可以验证 $\left(\mathsf{A}\frac{\psi_i}{\sqrt{\lambda_i}}\right)_{i\in\mathbb{N}}$ 是 $\mathcal{H}_2$ 的标准正交基： $\left\langle \mathsf{A}\frac{\psi_i}{\sqrt{\lambda_i}}, \mathsf{A}\frac{\psi_j}{\sqrt{\lambda_j}}\right\rangle = (\lambda_i\lambda_j)^{-1/2}\langle \psi_i, \mathsf{A}^*\mathsf{A}\psi_j\rangle = \sqrt{\frac{\lambda_j}{\lambda_i}}\langle \psi_i, \psi_j\rangle = \delta_{ij}$。 ■

附录 B. 方差项的界

我们开始 首先回顾通常的集中不等式，这将有助于处理方差项。

B.1. 集中不等式。 我们首先回顾一些关于随机向量和算子之和的集中不等式。

命题 B.1 (随机向量之和的伯恩斯坦不等式). 设 $z_1, \ldots, z_n$ 是可分希尔伯特空间 $\mathcal{H}$ 中一列独立同分布的随机元素。假设 $\mathbb{E}\|z_1\| < +\infty$，并注意 $\mu = \mathbb{E}z_1$。设 $\sigma, L \geqslant 0$ 满足：

$$\forall p \geqslant 2, \qquad \mathbb{E}\left\|z_1 - \mu\right\|_{\mathcal{H}}^p \leqslant \frac{1}{2} p! \sigma^2 L^{p-2}.$$

那么，对于任意 $\delta \in (0, 1]$,

$$\left\| \frac{1}{n} \sum_{i=1}^{n} z_i - \mu \right\|_{\mathcal{H}} \leqslant \frac{2L \log(2/\delta)}{n} + \sqrt{\frac{2\sigma^2 \log(2/\delta)}{n}}, \tag{18}$$

其概率至少为 $1 - \delta$。

证明。 这是对 [Yur95]中定理 3.3.4 的重新表述。 ■

命题 B.2 （随机算子之和的伯恩斯坦不等式）。 设 $\mathcal{H}$ 为可分希尔伯特空间，并设 X_1、.....、 X_n 是 $\mathcal{H}$上的一列独立同分布的自伴随机算子。假设 $\mathbb{E}$（X_i） $= 0$，且存在 $T > 0$ 和 S 一个正的迹类算子，使得 $\|X_i\| \leqslant T$几乎处处成立，并且对任意 $i \in \{1, \ldots, n\}$都有 $\mathbb{E}X_i^2 \preccurlyeq S$。那么，对任意 $\delta \in$ $(0, 1]$,，以下不等式成立：

$$\left\| \frac{1}{n} \sum_{i=1}^{n} X_i \right\| \leqslant \frac{2T\beta}{3n} + \sqrt{\frac{2\|S\|\beta}{n}}, \tag{19}$$

其概率至少为 $1 - \delta$，其中 $\beta = \log {}^{2\mathrm{Tr}}\|S\|\delta^{S}$。 ——

证明。 该定理是对 [Tro12] 中定理 7.3.1 的重新表述，并通过 [Sta17]中第 4 节的技术将其推广到可分希尔伯特空间情形。 ■

B.2. 算子界。

引理 B.3。 在假设 *2* 和 *3*下， Σ以及 L 均为迹类算子。

证明。 我们仅证明 L 的结果， Σ 的证明类似。考虑一个$\mathcal{H}$的标准正交基 $(\phi_i)_{i\in\mathbb{N}}$。由于 L 是正自伴算子，

$$\begin{aligned}
\operatorname{tr} \mathsf{L} &= \sum_{i=1}^{\infty} \langle \mathsf{L}\phi_i, \phi_i \rangle = \sum_{i=1}^{\infty} \mathbb{E}_\mu \left[\sum_{j=1}^{d} \langle \partial_j K_x, \phi_i \rangle^2 \right] = \mathbb{E}_\mu \left[\sum_{i=1}^{\infty} \sum_{j=1}^{d} \langle \partial_j K_x, \phi_i \rangle^2 \right] \\
&= \mathbb{E}_\mu \left[\sum_{j=1}^{d} \|\partial_j K_x\|^2 \right] \leqslant \mathcal{K}_d.
\end{aligned}$$

因此，L 是一个迹类算子。 ■

B.3. **命题 4.1的证明：方差界。** 这里回顾一下我们想要控制的估计量的方差表达式：

$$\begin{aligned}
\text{Variance} &:= \left\|\widehat{\mathsf{L}}_\lambda^{-1/2}\widehat{\Sigma}\widehat{\mathsf{L}}_\lambda^{-1/2} - \mathsf{L}_\lambda^{-1/2}\Sigma\mathsf{L}_\lambda^{-1/2}\right\| \\
&= \left\|\widehat{\mathsf{L}}_\lambda^{-1/2}\widehat{\Sigma}\widehat{\mathsf{L}}_\lambda^{-1/2} - \widehat{\mathsf{L}}_\lambda^{-1/2}\Sigma\widehat{\mathsf{L}}_\lambda^{-1/2}\right\| + \left\|\widehat{\mathsf{L}}_\lambda^{-1/2}\Sigma\widehat{\mathsf{L}}_\lambda^{-1/2} - \mathsf{L}_\lambda^{-1/2}\Sigma\mathsf{L}_\lambda^{-1/2}\right\| \\
&\leqslant \left\|\widehat{\mathsf{L}}_\lambda^{-1/2}(\widehat{\Sigma}-\Sigma)\widehat{\mathsf{L}}_\lambda^{-1/2}\right\| + \left\|\widehat{\mathsf{L}}_\lambda^{-1/2}\Sigma\widehat{\mathsf{L}}_\lambda^{-1/2} - \mathsf{L}_\lambda^{-1/2}\Sigma\widehat{\mathsf{L}}_\lambda^{-1/2}\right\| + \left\|\mathsf{L}_\lambda^{-1/2}\Sigma\widehat{\mathsf{L}}_\lambda^{-1/2} - \mathsf{L}_\lambda^{-1/2}\Sigma\mathsf{L}_\lambda^{-1/2}\right\| \\
&\leqslant \underbrace{\left\|\widehat{\mathsf{L}}_\lambda^{-1/2}(\widehat{\Sigma}-\Sigma)\widehat{\mathsf{L}}_\lambda^{-1/2}\right\|}_{\text{Lemma B.5}} + \underbrace{\left\|(\widehat{\mathsf{L}}_\lambda^{-1/2} - \mathsf{L}_\lambda^{-1/2})\Sigma\widehat{\mathsf{L}}_\lambda^{-1/2}\right\|}_{\text{Lemma B.6}} + \underbrace{\left\|\mathsf{L}_\lambda^{-1/2}\Sigma(\widehat{\mathsf{L}}_\lambda^{-1/2} - \mathsf{L}_\lambda^{-1/2})\right\|}_{\text{Lemma B.6 as well}}.
\end{aligned}$$

以下数量对于本节中的估计很有用：

$$\mathcal{N}_\infty(\lambda) = \sup_{x\in\mathrm{supp}(\mu)} \left\|\mathsf{L}_\lambda^{-1/2}K_x\right\|_{\mathcal{H}}^2, \text{ and } \quad \mathcal{F}_\infty(\lambda) = \sup_{x\in\mathrm{supp}(\mu)} \left\|\mathsf{L}_\lambda^{-1/2}\nabla K_x\right\|_{\mathcal{H}}^2.$$

注意，在假设 3 下，$\mathcal{N}_\infty(\lambda) \leqslant \frac{\mathcal{K}}{\lambda}$ 以及 $\mathcal{F}_\infty(\lambda) \leqslant \frac{\mathcal{K}_d}{\lambda}$。还要注意，在对 L 的谱作出更精细的假设时，后一个界相对于 λ 的依赖关系可能会更好。下面给出总体结果，然后再证明所有辅助引理：

引理 B.4。 对于任意 $0 < \lambda < \|\mathsf{L}\|$、$n \geqslant 15\mathcal{F}_\infty(\lambda)\log\frac{4\mathrm{Tr}\mathsf{L}}{\lambda\delta}$以及任意 $\delta \in (0,\ 1/2]$,

$$\begin{aligned}
\left\|\widehat{\mathsf{L}}_\lambda^{-1/2}\widehat{\Sigma}\widehat{\mathsf{L}}_\lambda^{-1/2} - \mathsf{L}_\lambda^{-1/2}\Sigma\mathsf{L}_\lambda^{-1/2}\right\| &\leqslant \frac{4\mathcal{N}_\infty(\lambda)\log\frac{2\mathcal{P}\mathrm{Tr}\Sigma}{\lambda\delta}}{3n} + \left[\frac{2\,\mathcal{P}\,\mathcal{N}_\infty(\lambda)\log\frac{4\mathcal{P}\mathrm{Tr}\Sigma}{\lambda\delta}}{n}\right]^{1/2} \\
&\quad + 4\left(\mathcal{P}\,\|\Sigma\|\right)^{1/2}\left(\frac{4\mathcal{F}_\infty(\lambda)\log\frac{4\,\mathrm{Tr}\mathsf{L}}{\lambda\delta}}{3n} + \sqrt{\frac{2\,\mathcal{F}_\infty(\lambda)\log\frac{4\,\mathrm{Tr}\mathsf{L}}{\lambda\delta}}{n}}\right),
\end{aligned}$$

以至少 $1-2\delta$ 的概率。

B.3.1。 第一项的界。

引理 B.5。 对于任意 $\lambda > 0$，以及任意 $\delta \in (0, 1]$,

$$\left\|\mathsf{L}_\lambda^{-1/2}(\widehat{\Sigma}-\Sigma)\mathsf{L}_\lambda^{-1/2}\right\| \leqslant \frac{4\mathcal{N}_\infty(\lambda)\log\frac{2\mathcal{P}\mathrm{Tr}\Sigma}{\lambda\delta}}{3n} + \left[\frac{2\,\mathcal{P}\,\mathcal{N}_\infty(\lambda)\log\frac{4\mathcal{P}\mathrm{Tr}\Sigma}{\lambda\delta}}{n}\right]^{1/2},$$

其概率至少为 $1-\delta$。

引理*B.5*的证明。 我们对算子 $\mathsf{L}_\lambda^{-1/2}\widehat{\Sigma}\mathsf{L}_\lambda^{-1/2}$ 应用某个集中不等式，其均值恰好为$\mathsf{L}_\lambda^{-1/2}\Sigma\mathsf{L}_\lambda^{-1/2}$。计算如下：

$$\begin{aligned}
\left\|\mathsf{L}_\lambda^{-1/2}(\widehat{\Sigma}-\Sigma)\mathsf{L}_\lambda^{-1/2}\right\| &= \left\|\mathsf{L}_\lambda^{-1/2}\widehat{\Sigma}\mathsf{L}_\lambda^{-1/2} - \mathsf{L}_\lambda^{-1/2}\Sigma\mathsf{L}_\lambda^{-1/2}\right\| \\
&= \left\|\frac{1}{n}\sum_{i=1}^n\left[(\mathsf{L}_\lambda^{-1/2}K_{x_i})\otimes(\mathsf{L}_\lambda^{-1/2}K_{x_i}) - \mathsf{L}_\lambda^{-1/2}\Sigma\mathsf{L}_\lambda^{-1/2}\right]\right\|.
\end{aligned}$$

我们使用命题 B.2。为此，我们对 $i \in [\![1,\ n]\!]$ 进行界定：

$$\begin{aligned}
\left\|(\mathsf{L}_\lambda^{-1/2}K_{x_i})\otimes(\mathsf{L}_\lambda^{-1/2}K_{x_i}) - \mathsf{L}_\lambda^{-1/2}\Sigma\mathsf{L}_\lambda^{-1/2}\right\| &\leqslant \left\|\mathsf{L}_\lambda^{-1/2}K_{x_i}\right\|_{\mathcal{H}}^2 + \left\|\mathsf{L}_\lambda^{-1/2}\Sigma\mathsf{L}_\lambda^{-1/2}\right\| \\
&\leqslant 2\mathcal{N}_\infty(\lambda),
\end{aligned}$$

并且，对于二阶矩，

$$\begin{aligned}
&\mathbb{E}\left((\mathsf{L}_\lambda^{-1/2}K_{x_i})\otimes(\mathsf{L}_\lambda^{-1/2}K_{x_i})-\mathsf{L}_\lambda^{-1/2}\Sigma\mathsf{L}_\lambda^{-1/2}\right)^2\\
&\quad=\mathbb{E}\left[\left\|\mathsf{L}_\lambda^{-1/2}K_{x_i}\right\|_{\mathcal{H}}^2(\mathsf{L}_\lambda^{-1/2}K_{x_i})\otimes(\mathsf{L}_\lambda^{-1/2}K_{x_i})\right]-\mathsf{L}_\lambda^{-1/2}\Sigma\mathsf{L}_\lambda^{-1}\Sigma\mathsf{L}_\lambda^{-1/2}\\
&\quad\preccurlyeq\mathcal{N}_\infty(\lambda)\mathsf{L}_\lambda^{-1/2}\Sigma\mathsf{L}_\lambda^{-1/2}.
\end{aligned}$$

我们通过对常数进行一些估计来完成证明： $\beta=\log\frac{2\mathrm{Tr}(\Sigma\mathsf{L}_\lambda^{-1})}{\left\|\mathsf{L}_\lambda^{-1/2}\Sigma\mathsf{L}_\lambda^{-1/2}\right\|\delta}$。为此，我们注意到，得益于 [PV20],中的命题 16

$$\left\|\mathsf{L}_\lambda^{-1/2}\Sigma\mathsf{L}_\lambda^{-1/2}\right\|=(\mathcal{P}_\mu^\lambda)^{-1}\geqslant\mathcal{P}^{-1},$$

并利用 $\mathrm{Tr}\Sigma\mathsf{L}_\lambda^{-1}\leqslant\lambda^{-1}\mathrm{Tr}\Sigma$，可得 $\beta\leqslant\log\frac{2\mathcal{P}\mathrm{Tr}\Sigma}{\lambda\delta}$。因此，

$$\begin{aligned}
&\left\|\frac{1}{n}\sum_{i=1}^n\left[(\mathsf{L}_\lambda^{-1/2}K_{x_i})\otimes(\mathsf{L}_\lambda^{-1/2}K_{x_i})-\mathsf{L}_\lambda^{-1/2}\Sigma\mathsf{L}_\lambda^{-1/2}\right]\right\|\\
&\qquad\leqslant\frac{4\mathcal{N}_\infty(\lambda)\log\frac{2\mathcal{P}\mathrm{Tr}\Sigma}{\lambda\delta}}{3n}+\left[\frac{2\,\mathcal{P}_\mu^\lambda\,\mathcal{N}_\infty(\lambda)\log\frac{2\mathcal{P}\mathrm{Tr}\Sigma}{\lambda\delta}}{n}\right]^{1/2}.
\end{aligned}$$

引理B.5的证明至此完成。 ■

B.3.2. 第二项的界。 这里，我们要对项 $\left\|(\widehat{\mathsf{L}}_\lambda^{-1/2}-\mathsf{L}_\lambda^{-1/2})\Sigma\widehat{\mathsf{L}}_\lambda^{-1/2}\right\|$ 进行界定。我们再来深入分析一下。

$$\begin{aligned}
\left\|(\widehat{\mathsf{L}}_\lambda^{-1/2}-\mathsf{L}_\lambda^{-1/2})\Sigma\widehat{\mathsf{L}}_\lambda^{-1/2}\right\|&=\left\|\widehat{\mathsf{L}}_\lambda^{1/2}(\widehat{\mathsf{L}}_\lambda^{-1}-\mathsf{L}_\lambda^{-1})\mathsf{L}_\lambda^{1/2}\Sigma\widehat{\mathsf{L}}_\lambda^{-1/2}\right\|\\
&=\left\|\widehat{\mathsf{L}}_\lambda^{-1/2}(\widehat{\mathsf{L}}_\lambda-\mathsf{L}_\lambda)\mathsf{L}_\lambda^{-1/2}\Sigma\widehat{\mathsf{L}}_\lambda^{-1/2}\right\|\\
&=\left\|\widehat{\mathsf{L}}_\lambda^{-1/2}\mathsf{L}_\lambda^{1/2}\mathsf{L}_\lambda^{-1/2}(\widehat{\mathsf{L}}_\lambda-\mathsf{L}_\lambda)\mathsf{L}_\lambda^{-1/2}\Sigma^{1/2}\Sigma^{1/2}\mathsf{L}_\lambda^{-1/2}\mathsf{L}_\lambda^{1/2}\widehat{\mathsf{L}}_\lambda^{-1/2}\right\|\\
&\leq\left\|\widehat{\mathsf{L}}_\lambda^{-1/2}\mathsf{L}_\lambda^{1/2}\right\|\left\|\mathsf{L}_\lambda^{-1/2}(\widehat{\mathsf{L}}_\lambda-\mathsf{L}_\lambda)\mathsf{L}_\lambda^{-1/2}\right\|\left\|\Sigma^{1/2}\right\|\left\|\Sigma^{1/2}\mathsf{L}_\lambda^{-1/2}\right\|\left\|\widehat{\mathsf{L}}_\lambda^{-1/2}\mathsf{L}_\lambda^{1/2}\right\|\\
&\leq(\mathcal{P}\,\|\Sigma\|)^{1/2}\underbrace{\left\|\widehat{\mathsf{L}}_\lambda^{-1/2}\mathsf{L}_\lambda^{1/2}\right\|^2}_{\text{Lemma B.8}}\underbrace{\left\|\mathsf{L}_\lambda^{-1/2}(\widehat{\mathsf{L}}_\lambda-\mathsf{L}_\lambda)\mathsf{L}_\lambda^{-1/2}\right\|}_{\text{Lemma B.7}}.
\end{aligned}$$

因此，我们可以表述本小节的主要结果：

引理 B.6。 对于任意 $0<\lambda<\|\mathsf{L}\|$， $\delta\in(0,\,1)$，以及 $n\geqslant15\mathcal{F}_\infty(\lambda)\log\frac{4\mathrm{Tr}\mathsf{L}}{\lambda\delta}$，以下结论以至少$1-\delta$的概率成立：

$$\left\|(\widehat{\mathsf{L}}_\lambda^{-1/2}-\mathsf{L}_\lambda^{-1/2})\Sigma\widehat{\mathsf{L}}_\lambda^{-1/2}\right\|\leq2\,(\mathcal{P}\,\|\Sigma\|)^{1/2}\left(\frac{4\mathcal{F}_\infty(\lambda)\log\frac{4\,\mathrm{Tr}\mathsf{L}}{\lambda\delta}}{3n}+\sqrt{\frac{2\,\mathcal{F}_\infty(\lambda)\log\frac{4\,\mathrm{Tr}\mathsf{L}}{\lambda\delta}}{n}}\right),$$

以至少 $1-\delta$ 的概率。

现在我们来证明这些辅助引理。

引理 B.7。 对于任意 $0<\lambda<\|\mathsf{L}\|$ 以及任意 $\delta\in(0,1]$,

$$\left\|\mathsf{L}_\lambda^{-1/2}(\widehat{\mathsf{L}}-\mathsf{L})\mathsf{L}_\lambda^{-1/2}\right\|\leqslant\frac{4\mathcal{F}_\infty(\lambda)\log\frac{4\,\mathrm{Tr}\mathsf{L}}{\lambda\delta}}{3n}+\sqrt{\frac{2\,\mathcal{F}_\infty(\lambda)\log\frac{4\,\mathrm{Tr}\mathsf{L}}{\lambda\delta}}{n}},$$

其概率至少为 $1-\delta$。

引理的证明*B.7*。 与引理 B.5 的证明一样，我们希望对算子 $\mathsf{L}_\lambda^{-1/2}\widehat{\Delta}\mathsf{L}_\lambda^{-1/2}$ 应用某个集中不等式，其均值恰好为 $\mathsf{L}_\lambda^{-1/2}\Delta\mathsf{L}_\lambda^{-1/2}$。证明几乎与引理 B.5 相同。我们先写出

$$\begin{aligned}\left\|\mathsf{L}_\lambda^{-1/2}(\widehat{\mathsf{L}}-\mathsf{L})\mathsf{L}_\lambda^{-1/2}\right\| &= \left\|\mathsf{L}_\lambda^{-1/2}\widehat{\mathsf{L}}\mathsf{L}_\lambda^{-1/2}-\mathsf{L}_\lambda^{-1/2}\mathsf{L}\mathsf{L}_\lambda^{-1/2}\right\| \\ &= \left\|\frac{1}{n}\sum_{i=1}^{n}\left[(\mathsf{L}_\lambda^{-1/2}\nabla K_{x_i})\otimes(\mathsf{L}_\lambda^{-1/2}\nabla K_{x_i})-\mathsf{L}_\lambda^{-1/2}\Delta\mathsf{L}_\lambda^{-1/2}\right]\right\|.\end{aligned}$$

为了使用命题 B.2，我们对 $i\in[\![1,\ n]\!]$，

$$\begin{aligned}\left\|(\mathsf{L}_\lambda^{-1/2}\nabla K_{x_i})\otimes(\mathsf{L}_\lambda^{-1/2}\nabla K_{x_i})-\mathsf{L}_\lambda^{-1/2}\mathsf{L}\mathsf{L}_\lambda^{-1/2}\right\| &\leqslant \left\|\mathsf{L}_\lambda^{-1/2}\nabla K_{x_i}\right\|_{\mathcal{H}}^2+\left\|\mathsf{L}_\lambda^{-1/2}\mathsf{L}\mathsf{L}_\lambda^{-1/2}\right\| \\ &\leqslant 2\mathcal{F}_\infty(\lambda),\end{aligned}$$

以及二阶矩，

$$\begin{aligned}&\mathbb{E}\left[\left((\mathsf{L}_\lambda^{-1/2}\nabla K_{x_i})\otimes(\mathsf{L}_\lambda^{-1/2}\nabla K_{x_i})-\mathsf{L}_\lambda^{-1/2}\mathsf{L}\mathsf{L}_\lambda^{-1/2}\right)^2\right] \\ &=\mathbb{E}\left[\left\|\mathsf{L}_\lambda^{-1/2}\nabla K_{x_i}\right\|_{\mathcal{H}}^2(\mathsf{L}_\lambda^{-1/2}\nabla K_{x_i})\otimes(\mathsf{L}_\lambda^{-1/2}\nabla K_{x_i})\right]-\mathsf{L}_\lambda^{-1/2}\mathsf{L}\mathsf{L}_\lambda^{-1}\mathsf{L}\mathsf{L}_\lambda^{-1/2} \\ &\preccurlyeq \mathcal{F}_\infty(\lambda)\mathsf{L}_\lambda^{-1/2}\mathsf{L}\mathsf{L}_\lambda^{-1/2}.\end{aligned}$$

最后，我们对 $\beta=\log\frac{2\mathrm{Tr}(\mathsf{L}\mathsf{L}_\lambda^{-1})\left\|\mathsf{L}_\lambda^{-1}\mathsf{L}\right\|\delta}{4\mathrm{Tr}\mathsf{L}}$ 进行一些估计。由于 $\mathrm{Tr}(\mathsf{L}\mathsf{L}_\lambda^{-1})\leqslant\lambda^{-1}\mathrm{Tr}\mathsf{L}$ 且对于 $\lambda\leqslant\|\mathsf{L}\|$，$\left\|\mathsf{L}_\lambda^{-1}\mathsf{L}\right\|\geqslant 1/2$，因此 $\beta\leqslant\log\frac{\lambda\delta}{4\mathrm{Tr}\mathsf{L}}$。由此可得结论（19）。 ■

引理 B.8 (算子界)。 对于任意 $\lambda>0$, $\delta\in(0,1)$，以及 $n\geqslant 15\mathcal{F}_\infty(\lambda)\log\frac{4\mathrm{Tr}\mathsf{L}}{\lambda\delta}$，以至少 $1-\delta$ 的概率有：

$$\left\|\widehat{\mathsf{L}}_\lambda^{-1/2}\mathsf{L}_\lambda^{1/2}\right\|^2\leqslant 2,$$

该结果的证明依赖于以下引理（参见 [RR17，命题 8] 的证明）。

引理 B.9。 设 $\mathcal{H}$ 为可分希尔伯特空间，A 和 B 为作用于 $\mathcal{H}$ 和 $\lambda>0$上的两个有界自伴正线性算子。则

$$\left\|(A+\lambda I)^{-1/2}(B+\lambda I)^{1/2}\right\|\leqslant(1-\beta)^{-1/2},$$

其中 $\beta=\lambda_{\max}\left((B+\lambda I)^{-1/2}(B-A)(B+\lambda I)^{-1/2}\right)<1$，其中 $\lambda_{\max}(O)$ 是自伴算子 O的最大特征值。

现在可以写出引理B.8的证明。

引理 *B.8*的证明。 根据引理 B.9，我们看到

$$\left\|\widehat{\mathsf{L}}_\lambda^{-1/2}\mathsf{L}_\lambda^{1/2}\right\|^2\leqslant\left(1-\lambda_{\max}\left(\mathsf{L}_\lambda^{-1/2}(\widehat{\mathsf{L}}-\mathsf{L})\mathsf{L}_\lambda^{-1/2}\right)\right)^{-1},$$

并且由于 $\left\|\mathsf{L}_\lambda^{-1/2}(\widehat{\mathsf{L}}-\mathsf{L})\mathsf{L}_\lambda^{-1/2}\right\|<1$，我们有：

$$\left\|\widehat{\mathsf{L}}_\lambda^{-1/2}\mathsf{L}_\lambda^{1/2}\right\|^2\leqslant\left(1-\left\|\mathsf{L}_\lambda^{-1/2}(\widehat{\mathsf{L}}-\mathsf{L})\mathsf{L}_\lambda^{-1/2}\right\|\right)^{-1}.$$

然后，我们可以应用引理 B.7 中的界，得到：若 λ满足 $\frac{4\mathcal{F}_\infty(\lambda)\log\frac{4\mathrm{Tr}\mathsf{L}}{\lambda\delta}}{3n}+\sqrt{\frac{2\mathcal{F}_\infty\mathrm{Tr}\mathsf{L}(\lambda)\log\frac{4\mathrm{Tr}\mathsf{L}}{\lambda\delta}}{n}}\leqslant\frac{1}{2}$，则以概率 $1-\delta$有 $\left\|\widehat{\mathsf{L}}_\lambda^{-1/2}\mathsf{L}_\lambda^{1/2}\right\|^2\leqslant 2$。当 $n\geqslant 15\mathcal{F}_\infty(\lambda)\log\frac{4\mathrm{Tr}\mathsf{L}}{\lambda\delta}$时，$\lambda$的条件得到满足。 ■

B.3.3. 第三项（也是最后一项）的界。 这里，我们要对项 $\left\|\mathsf{L}_\lambda^{-1/2}\Sigma(\widehat{\mathsf{L}}_\lambda^{-1/2}-\mathsf{L}_\lambda^{-1/2})\right\|$进行界定。我们采用与前面相同的技巧。

$$\begin{aligned}\left\|\mathsf{L}_\lambda^{-1/2}\Sigma(\widehat{\mathsf{L}}_\lambda^{-1/2}-\mathsf{L}_\lambda^{-1/2})\right\| &= \left\|\mathsf{L}_\lambda^{-1/2}\Sigma\mathsf{L}_\lambda^{-1/2}(\mathsf{L}-\widehat{\mathsf{L}})\widehat{\mathsf{L}}_\lambda^{-1/2}\right\| \\ &\leq (\mathcal{P}\,\|\Sigma\|)^{1/2}\left\|\widehat{\mathsf{L}}_\lambda^{-1/2}\mathsf{L}_\lambda^{1/2}\right\|^2\left\|\mathsf{L}_\lambda^{-1/2}(\widehat{\mathsf{L}}_\lambda-\mathsf{L}_\lambda)\mathsf{L}_\lambda^{-1/2}\right\|.\end{aligned}$$

因此，适用与引理 B.6 相同的界！

附录 C. 偏差项

C.1. **一致性证明：命题 4.2。** 为证明命题 4.2，我们首先需要一个关于算子范数收敛的一般结果。

引理 C.1. 设 $\mathcal{H}$ 是一个希尔伯特空间，并假设 $(A_n)_{n\geqslant 0}$ 是一个有界算子族，满足$\forall n\in\mathbb{N}$, $\|A_n\|\leqslant 1$ 和 $\forall f\in\mathcal{H}$, $A_nf\xrightarrow{n\to\infty}Af$。另设 B 是一个紧算子。那么，在算子范数意义下，

$$A_nBA_n^*\xrightarrow{n\to\infty}ABA^*.$$

证明。 设 $\varepsilon>0$。由于 B是紧算子，它可以由一个有限秩算子 $B_{n_\varepsilon}=\sum_{i=1}^{n_\varepsilon}b_i\langle f_i,\cdot\rangle g_i$近似，其中 $(f_i)_i$ 和 $(g_i)_i$ 是标准正交基，而 $(b_i)_i$ 是一个极限为零的非负数列（该算子的奇异值）。更具体地说， n_ε 的选取满足

$$\|B-B_{n_\varepsilon}\|\leqslant\frac{\varepsilon}{2}.$$

此外，固定 ε， $A_nB_{n_\varepsilon}A_n^*=\sum_{i=1}^{n_\varepsilon}b_i\langle A_nf_i,\cdot\rangle A_ng_i\underset{n\infty}{\longrightarrow}\sum_{i=1}^{n_\varepsilon}b_i\langle Af_i,\cdot\rangle Ag_i=AB_{n_\varepsilon}A^*$ 在算子范数下，因此，对于 $n\geqslant N_\varepsilon$，当 $N_\varepsilon\geqslant n_\varepsilon$ 足够大时，$\|A_nB_{n_\varepsilon}A_n^*-AB_{n_\varepsilon}A^*\|\leqslant\frac{\varepsilon}{2}$。最后，当 $\|A\|\leqslant 1$时，对于 $n\geqslant N_\varepsilon$，有

$$\begin{aligned}\|A_nB_{n_\varepsilon}A_n^*-ABA^*\| &\leqslant \|A_nB_{n_\varepsilon}A_n^*-AB_{n_\varepsilon}A^*\|+\|A(B_{n_\varepsilon}-B)A^*\| \\ &\leqslant \|A_nB_{n_\varepsilon}A_n^*-AB_{n_\varepsilon}A^*\|+\|B_{n_\varepsilon}-B\|\leqslant\varepsilon.\end{aligned}$$

This 证明了当 n 趋于无穷大时， $A_nBA_n^*$ 在算子范数下收敛于 ABA^* . ■

现在我们可以证明命题4.2。

命题 4.2 的证明。记 $B=\mathrm{L}^{-1/2}\Sigma\mathrm{L}^{-1/2}$ 和 $A_\lambda=\mathrm{L}_\lambda^{-1/2}\mathrm{L}^{1/2}$ 均定义在 $\mathcal{H}_0$ 上，则有 $\mathrm{L}_\lambda^{-1/2}\Sigma\mathrm{L}_\lambda^{-1/2}=A_\lambda BA_\lambda^*$，其中 B 为紧算子，且 $\|A_\lambda\|\leqslant 1$。此外，设 $(\phi_i)_{i\in\mathbb{N}}$ 是紧算子 L 的一组标准正交特征向量族，对应特征值为 $(\nu_i)_{i\in\mathbb{N}}$。那么，对任意$f\in\mathcal{H}_0$，我们可以写成：

$$A_\lambda f=\mathsf{L}_\lambda^{-1/2}\mathsf{L}^{1/2}f=\sum_{i=0}^{\infty}\sqrt{\frac{\nu_i}{\lambda+\nu_i}}\langle f,\phi_i\rangle_{\mathcal{H}}\,\phi_i\xrightarrow[\lambda\to 0]{}f.$$

因此，应用引理 C.1，我们可得 $\mathsf{L}_\lambda^{-1/2}\Sigma\mathsf{L}_\lambda^{-1/2}$ 在算子范数意义下收敛到 $\mathsf{L}^{-1/2}\Sigma\mathsf{L}^{-1/2}$。

■

C.2. **源条件下的快速收敛率：命题4.3。**

命题 *4.3* 的证明。为证明命题 4.3，我们只需根据以下不等式对偏差项进行界定。

$$\begin{aligned}\left\|\Pi^p\left(\mathsf{L}_\lambda^{-1/2}\Sigma\mathsf{L}_\lambda^{-1/2}-\mathsf{L}^{-1/2}\Sigma\mathsf{L}^{-1/2}\right)\Pi^p\right\| &\le \left\|\Pi^p\left(\mathsf{L}_\lambda^{-1/2}-\mathsf{L}^{-1/2}\right)\Sigma\mathsf{L}_\lambda^{-1/2}\Pi^p\right\| \\ &\quad+\left\|\Pi^p\mathsf{L}^{-1/2}\Sigma\left(\mathsf{L}_\lambda^{-1/2}-\mathsf{L}^{-1/2}\right)\Pi^p\right\| \\ &\le \left\|\Pi^p\mathsf{L}_\lambda^{-1/2}\mathsf{L}^{-1/2}\Sigma\mathsf{L}^{-1/2}\Pi^p\right\|+\left\|\Pi^p\mathsf{L}^{-1/2}\Sigma\mathsf{L}^{-1/2}\mathsf{L}_\lambda^{-1/2}\Pi^p\right\| \\ &\le \left\|\Pi^p\mathsf{L}^{-1/2}\mathsf{L}^{-1/2}\Sigma\mathsf{L}^{-1/2}\Pi^p\right\|+\left\|\Pi^p\mathsf{L}^{-1/2}\Sigma\mathsf{L}^{-1/2}\mathsf{L}^{-1/2}\Pi^p\right\| \\ &\le 2\mathcal{P}\left\|\Pi^p\mathsf{L}^{-1/2}\Pi^p\right\|,\end{aligned}$$

这最终证明了命题4.3。 ■

当然，如果对 $\mathcal{L}$的特征向量光滑程度先验条件更精细，即对类似于 $\left\|\Pi^p\mathsf{L}^{-\theta}\Pi^p\right\|$的控制，其中 $\theta\in[0,1]$ 可以推广我们所用的源条件，则可得到更细致的收敛速率 [PVRB18、BBG20、VPVF21]。