arXiv:2603.00971v1 [stat.ML] 1 Mar 2026

算子值核的随机特征：连接核方法与神经算子

Mike Nguyen
布伦瑞克工业大学

Nicole M¨ucke
布伦瑞克工业大学

摘要

本文研究了随机特征方法的泛化性质。我们的分析将关于 Tikhonov 正则化的已有结果拓展至一大类谱正则化技术，并进一步将研究框架推广到算子值核。这一统一框架使我们能够从神经切线核（NTK）的视角，对神经算子和神经网络进行严谨的理论分析。具体而言，它使我们能够确立最优学习率，并深入理解达到给定精度所需的神经元数量。此外，我们还确立了良设定情形和错设定情形下的极小极大速率；在错设定情形中，目标函数不属于再生核希尔伯特空间。这些结果进一步完善并深化了此前针对特定核算法的研究发现。

1 引言

算子学习已成为机器学习中的一种强大范式。它尤其适用于不确定性量化、逆问题和设计优化等领域的代理建模。这些应用的核心目标是近似潜在的非线性算子，例如偏微分方程（PDEs）的解算子，参见Kovachki 等人（2023a）。在这一领域应用最广泛的方法包括神经算子（NOs），它将经典神经网络推广到更广泛的算子学习任务，即学习（可能是）无限维函数空间之间的映射（Kovachki 等人，2023b；Li等人，2021c；Qin 等人，2024；Wangetal.，2021；Raissi etal.，2019；Li等人，2020；Sharma etal.，2024）。

尽管神经算子（NOs）在科学计算中取得了实际成功，但我们对其理论的理解仍然有限。现有研究主要聚焦于逼近性质（Huang 等人，2024；Schwab 和 Zech，2021；Marcati 和 Schwab，2023；Kovachki 等人，2021，2024），而泛化方面的结果相对较少（例如，参见Kim 和 Kang（2022）；Lara Benitez 等人（2024））。近期，神经切线核（NTK）理论领域取得了进展，Nguyen 和 M¨ucke （2024）为算子学习建立了极小极大速率。与此同时，梯度下降（GD）的动力学揭示了其与向量值再生核随机特征近似之间的联系，但此类方法目前尚无收敛保证。迄今为止，在这一设定下仅有针对 Tikhonov 正则化的分析（Lanthaler 和 Nelsen，2023）。本文通过推导一大类谱正则化方案（包括梯度下降）的极小极大速率，进一步为推导神经算子的泛化界奠定基础。

除了弥合这一理论差距之外，随机特征近似（RFA）本身也具有独立的研究价值。核方法在许多非参数统计应用中仍处于最先进水平，并为发展新的理论见解提供了优雅的框架（Lin 和 Cevher，2018；Lin 等人，2020a；Zhang等人，2024）。然而，这些方法的优势伴随着高昂的计算成本，使其难以应用于大规模数据集。经典的核化算法需要存储核格拉姆矩阵 $\mathrm{K} \in \mathrm{R}^{n\times n}$，其元素为：对于核函数 $K(\cdot,\cdot)$ 和数据点 u_i, u_j，矩阵元素为 $\mathrm{K}_{i,j} = K(u_i, u_j)$。这需要 $O(n^2)$ 的内存开销，计算时间最高可达 $O(n^3)$，其中n 表示数据集大小（Schoelkopf and Smola，2002）。

RFA 利用核的积分表示，并通过随机特征的有限和对其进行近似，从而降低这些成本。对于核岭回归（KRR），这会将内存和计算成本分别降至 $O(nM)$ 和 $O(nM^2)$，

29th人工智能与统计学会议（人工智能与统计学会议）2026年论文集，摩洛哥丹吉尔。PMLR：第 300 卷。
版权归作者所有，2026年。

分别为，其中 M 是随机特征的数量（Rudi 和 Rosasco，2016年）。对于梯度下降（GD），成本变为 $O(nMt)$，其中t 表示迭代次数；采用重球法或 Nesterov加速等加速方法还可以进一步降低成本，它们仅需 $\sqrt{t}$ 次迭代即可达到与 GD 相同的泛化误差（Pagliana and Rosasco，2019）。我们的分析还进一步表明，M 通常应大于 t，这解释了为什么随机特征方法与迭代正则化方案结合时往往效果最佳。这一观察凸显了在此类情形下提供严格理论保证的必要性。

对于核岭回归（KRR），需要多少随机特征才能达到最优收敛速率这一核心问题已得到广泛研究（Rahimi and Recht，2007年；Li等人，2021b；Zhen等人，2020年；Rudi 和 Rosasco，2016年；Lanthaler 和 Nelsen（2023））。Rahimi 和 Recht（2007年）首次针对实值核（rvk）情形确立了使用 $M = O(n)$ 个随机特征时的最优收敛率。随后，Rudi 和 Rosasco（2016年）将其改进为 $M = O(\sqrt{n}\log n)$，并由 Carratino等人（2019）进一步扩展至随机核岭回归。最近，Lanthaler 和 Nelsen（2023）去除了对数因子，证明在更一般的向量值核（vvk）设定下，$M = O(\sqrt{n})$ 已足够。表 1 对这些结果进行了详细比较。对于 M，这些结果适用于目标函数属于再生核希尔伯特空间的良设定情形。对于更广泛的光滑度类别，如果目标函数可能不属于再生核希尔伯特空间，需要多少随机特征这一问题仍未解决。

贡献。 我们研究向量值核和随机特征近似（RFA）的主要动机，是为神经算子（NOs）推导泛化保证。

据我们所知，基于向量值核的 RFA 此前主要在核岭回归（KRR）的情形下得到分析。在本工作中，我们开发了一个基于谱滤波的统一框架（Caponnetto 和 De Vito，2007年），为一大类采用显式或隐式正则化的学习方法提供最优收敛率。这包括梯度下降和加速技术，并将实值核的 KRR 结果（Rudi 和 Rosasco （2016年）以及向量值核的结果（Lanthaler 和 Nelsen （2023）作为特例涵盖在内。

我们的框架还适用于表示为积分核之和的核，尤其涵盖算子值神经正切核。这一扩展首次为超越核岭回归的向量值核随机特征方法提供了收敛速率，从而开辟了在神经切线核（NTK）情形下为神经算子提供严格统计保证的道路。

我们方法的一项关键优势在于，收敛率和达到最优性所需的随机特征数量均不受输入空间维度的影响。这使得这些结果可直接应用于神经算子，因为其输入是函数，而非有限维向量。同时，随机特征的数量仅随每个神经元组合输入表示的特征维度呈二次增长，从而为神经算子提供兼具极小极大最优统计率与计算效率的泛化保证。

表 1：达到阶为 M 的泛化误差所需的随机特征数量（ O）（$n^{-\frac{1}{2}}$）。最后一列指出已知最优速率的光滑度类别；关于 3.2、3.3 中 r、b >0 的含义，请参阅假设 [1]=Rahimiand Recht （2007年）、[2]=Rudi 和 Rosasco （2016年）、[3]= Lanthaler 和 Nelsen（2023）

参考文献	M	方法	光滑性
[1]	$O(n)$	核岭回归（rvk）	$r \in [0.5, 1]$
[2]	$O(\sqrt{n}$对数 $n)$	核岭回归 (rvk)	$r \in [0.5, 1]$
[3]	$O(\sqrt{n})$	核岭回归 (vvk)	$r = 0.5$
Our	$O(\sqrt{n}$对数 $n)$	Spec. (vvk)	$2r + b > 1$

本文其余部分安排如下。第 2 节介绍问题设定，回顾随机特征方法背景下的关键定义，并通过将其与神经算子联系起来，阐明这一框架的动机。第 3 节介绍并讨论我们的主要结果。第 4 节总结研究发现。附录 A 进一步介绍神经算子学习的相关细节，并提供支持理论结果的数值示例。所有证明均收录于附录 B。

记号。我们用 $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ 表示实可分希尔伯特空间 $\mathcal{H}_1, \mathcal{H}_2$ 之间的有界线性算子空间。记 $\mathcal{L}(\mathcal{H}, \mathcal{H}) = \mathcal{L}(\mathcal{H})$。对于 $\Gamma \in \mathcal{L}(\mathcal{H})$，我们用 Γ^* 表示伴随算子。若 $h \in \mathcal{H}$，则记作 $h \otimes h := \langle \cdot, h\rangle h$。对于紧算子 $\Gamma \in \mathcal{L}(\mathcal{H})$，其迹定义为 $\mathrm{tr}(\Gamma) = \sum_{k=1}^{\infty}\langle \Gamma e_k, e_k\rangle$，其中 $\{e_k\}_{k=1}^{\infty}$ 是 $\mathcal{H}$的任意标准正交基。我们用$\mathcal{F}(\mathcal{U}, \mathcal{V})$表示从$\mathcal{U}$到 $\mathcal{V}$的可测算子空间。记 $L^2(\mathcal{U}, \rho_{\mathcal{U}}) := L^2(\mathcal{U}, \rho_{\mathcal{U}}; \mathcal{V})$，表示配备范数 $\|f\|^2_{L^2(\rho_{\mathcal{U}})} := \int_{\mathcal{U}} \|f(u)\|^2_{\mathcal{V}}\, d\rho_{\mathcal{U}}(u)$ 的 L^2空间。令 $[n] := \{1, \ldots, n\}$，并将输出向量记为 $\mathbf{v} = (v_1, \ldots, v_n) \in \mathcal{V}^n$，其范数为 $\|\mathbf{v}\|^2_2 := \sum_{i=1}^{n} \|v_i\|^2_{\mathcal{V}}$。

2 数学框架

我们考虑一个输入空间 $(\mathcal{U}, \|\cdot\|_{\mathcal{U}})$，其中 $\mathcal{U}$是巴拿赫空间，以及一个输出空间 $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$，其中 $\mathcal{V}$ 是可分希尔伯特空间。这些假-

设有助于运用向量值核理论（Carmeli 等人，2005年，2008年）。数据空间由 $\mathcal{Z} = \mathcal{U} \times \mathcal{V}$给出，并配有未知分布 ρ。我们用 $\rho_{\mathcal{U}}$ 表示 $\mathcal{U}$上的边缘分布，并用 $\rho(\cdot \mid u)$ 表示 $\mathcal{V}$ 上给定 $u \in \mathcal{U}$的正则条件分布；参见Shao （2003年）。

给定一个可测算子 $G: \mathcal{U} \to \mathcal{V}$，我们进一步将期望风险定义为

$$\mathcal{E}(G) := \mathbb{E}[\ell(G(u), v)] , \tag{2.1}$$

其中，期望是相对于分布 ρ 和 $\ell: \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ + 取的，$\ell(v, v') = \frac{1}{2}\|v - v'\|_{\mathcal{V}}^2$是最小二乘损失。已知，在所有可测函数构成的集合上，$\mathcal{E}$ 的全局极小值由回归算子$G_\rho(u) = \int_{\mathcal{V}} v\rho(dv|u)$给出。

我们考虑标准的统计学习设定：给定数据 $(u_j, v_j)_{j=1}^n$，这些数据是相对于 ρ 在$\mathcal{U} \times \mathcal{V}$上独立同分布采样得到的。

2.1 动机：神经算子的泛化界

浅层神经算子。为了将浅层神经算子（NOs）与向量值核联系起来，我们有必要回顾其定义，并说明其训练动态如何产生神经切线核（NTK）。根据 Nguyen 和 M¨ucke（2024），双层神经算子的类别定义如下。令$\mathcal{U}$ 表示函数输入空间，它将测度空间 $(\mathcal{X}, \rho_x)$ 映射到 $\mathcal{Y} \subset \mathbb{R}^{d_y}$；令 $\mathcal{V}$表示目标函数空间，其中包含将 $(\mathcal{X}, \rho_x)$ 映射到 $\mathcal{Y} \subset \mathbb{R}$ 的可测函数。对于宽度为 $M \in \mathbb{N}$ 的网络，令 $\sigma: \mathbb{R} \to \mathbb{R}$ 为逐点作用的激活函数，并令 $A: \mathcal{U} \to \mathcal{F}(\mathcal{X}, \mathbb{R}^{d_k})$ 为连续算子。于是，我们将浅层神经算子定义为

$$\begin{aligned}\mathcal{F}_M := \Big\{ G_\theta : \mathcal{U} \to \mathcal{V} \;\Big|\; & G_\theta(u)(x) \\ & = \frac{\langle a,\ \sigma(B_1 A(u)(x) + B_2 u(x) + B_3 c(x)) \rangle}{\sqrt{M}}, \\ & \theta = (a, B_1, B_2, B_3) \\ & \in \mathbb{R}^M \times \mathbb{R}^{M \times d_k} \times \mathbb{R}^{M \times d_y} \times \mathbb{R}^{M \times d_b}, \\ & c : \mathcal{X} \to \mathbb{R}^{d_b} \Big\}.\end{aligned}$$

其中，$\langle \cdot, \cdot \rangle$ 表示欧几里得内积。我们将所有参数汇总为 $\theta = (a, B_1, B_2, B_3) = \tilde{}(a, B)$，其中 $B = (B_1, B_2, B_3) \in \mathbb{R}^{M \times d}$，特征维度为 $d := d_k + d_y + d_b$。目标是在集合 $\mathcal{F}_M$上最小化期望风险（2.1），即

$$\min_{G_\theta \in \mathcal{F}_M} \mathbb{E}[\ell(G_\theta(u), v)] .$$

通过神经切线核建立与向量值核的联系。核方法与标准神经网络之间的联系通过神经切线核（NTK）建立，参见 Jacot 等人(2018）；Lee 等人(2019）。对于全连接神经网络，已知在梯度下降训练下，且处于无限宽极限时，网络函数会在初始化点附近线性化，其动态由神经切线核支配。完全类似地，浅层神经算子的训练动态对应于由向量值核诱导的再生核希尔伯特空间中的核梯度下降。我们为神经算子定义神经切线核特征映射

$$\Phi^M : \mathcal{U} \to \mathcal{F}(L^2(\mathcal{X}, \rho_x), \Theta), \qquad \Phi^M(u) := \Phi_u^M,$$

采用随机初始化 $\theta_0 \in \Theta$，并且

$$\Phi_u^M(v) := \nabla_\theta \langle G_{\theta_0}(u), v \rangle_{L^2(\rho_x)} , \qquad v \in L^2(\mathcal{X}, \rho_x).$$

对于任意 u，$\tilde{u} \in \mathcal{U}$，向量值核 $K_M: L^2(\mathcal{X}, \rho_x) \to L^2(\mathcal{X}, \rho_x)$定义为

$$K_M(u, \tilde{u}) := (\Phi_u^M)^* \Phi_{\tilde{u}}^M$$

并展开为

$$\begin{aligned}K_M(u, \tilde{u}) = & \frac{1}{M} \sum_{m=1}^{M} \psi_m(u) \otimes \psi_m(\tilde{u}) \\ & + \frac{1}{M} \sum_{m=1}^{M} \sum_{j=1}^{\tilde{d}} \psi'_{m,j}(u) \otimes \psi'_{m,j}(\tilde{u}).\end{aligned}$$

这里，对于 $u \in \mathcal{U}$ 和 $x \in \mathcal{X}$，我们设

$$\begin{aligned}J(u)(x) &:= \big(A(u)(x),\, u(x),\, c(x)\big)^\top, \\ \psi_m(u) &:= \sigma\big(\langle b_m^{(0)}, J(u) \rangle\big), \\ \psi'_{m,j}(u) &:= \sigma'\big(\langle b_m^{(0)}, J(u) \rangle\big) J(u)^{(j)}\end{aligned}$$

并将初始权重向量的集合记为 $B^{(0)} = (b_1^{(0)}, \ldots\ldots, b_M^{(0)})^\top \in \mathbb{R}^{M \times \tilde{d}}$，参见Nguyen 和 M¨ucke (2024)。

与随机特征近似的关系。 K_M 的表示表明，浅层神经算子的神经切线核具有随机特征近似的形式。事实上，每一项$\psi_m(u) \otimes \psi_m(\tilde{u})$和$\psi'_{m,\,j}(u) \otimes \psi'_{m,\,j}(\tilde{u})$都构成一个非线性随机特征，并通过随机初始化 $b_m^{(0)}$进行采样。因此，K_M 是极限核的蒙特卡洛近似

$$K(u, \tilde{u}) := \mathbb{E}_{\theta_0}\Big[\psi(u) \otimes \psi(\tilde{u}) + \sum_{j=1}^{\tilde{d}} \psi'_j(u) \otimes \psi'_j(\tilde{u})\Big], \tag{2.2}$$

其中 $\psi(u) = \sigma\big(\langle b^{(0)},\ J(u) \rangle\big)$，$\psi'_j(u) = \sigma'\big(\langle b^{(0)},\ J(u) \rangle\big) J(u)^{(j)}$，期望是相对于 θ_0 的初始化分布而言的，参见 NguyenandM¨ucke (2024)，命题 2.3。

因此，在神经切线核区域使用梯度下降训练浅层神经算子，等价于在与 K 相关的向量值再生核希尔伯特空间中执行核梯度下降，并将 K_M 作为随机特征近似。这一解释建立了神经算子与向量值核的随机特征统计理论之间的桥梁。

泛化界。 假设一个神经算子通过经验风险最小化，在类别 $\mathcal{F}_M$ 上使用梯度下降进行训练，并令 θ_t 表示经过 $t \in \mathbb{N}$ 次迭代后的参数更新。相应神经算子 $G_{\theta_t} \in \mathcal{F}_M$ 的超额风险可分解为

$$\begin{aligned} &\|G_{\theta_t} - G_\rho\|^2_{L^2(\rho_\mathcal{U})} \\ &\lesssim \|G_{\theta_t} - F_t^M\|^2_{L^2(\rho_\mathcal{U})} + \|F_t^M - G_\rho\|^2_{L^2(\rho_\mathcal{U})}, \end{aligned} \tag{2.3}$$

其中， F_t^M 表示与向量值核 K_M 相关联、并采用相同梯度下降动态训练的随机特征估计量。第一项刻画有限宽度神经算子与基于神经切线核的随机特征近似之间的差异，第二项则衡量随机特征方法本身的泛化误差。**本文的核心目标是为第二项建立最优学习保证，并将分析拓展至更广泛的正则化方案与核**。

2.2 核方法、随机特征与正则化：一般设定

向量值核。 设 $K\colon \mathcal{U}\times\mathcal{U} \longrightarrow \mathcal{L}(\mathcal{V})$ 是一个正定型再生 $\mathcal{V}$-值核，其中 $\mathcal{L}(\mathcal{V})$ 表示 $\mathcal{V}$ 上有界线性算子所构成的巴拿赫空间。对于 $u \in \mathcal{U}$，记 $K_u : \mathcal{V} \to \mathcal{F}(\mathcal{U}, \mathcal{V})$ 为将 $v \in \mathcal{V}$ 映射到函数 $K_u v \in \mathcal{F}(\mathcal{U}, \mathcal{V})$ 的算子，其定义为

$$(K_u v)(\tilde{u}) = K(\tilde{u}, u)v, \qquad \tilde{u} \in \mathcal{U}. \tag{2.4}$$

我们用 $\mathcal{H}$ 表示相应的唯一 $\mathcal{V}$值再生核希尔伯特空间，其定义在 $\mathcal{U}$上，并可连续包含于 $\mathcal{F}(\mathcal{U}, \mathcal{V})$ 中。我们假设 K 是二阶有界的，这意味着包含 $S: \mathcal{H} \hookrightarrow L^2(\mathcal{U}, \rho_\mathcal{U}; \mathcal{V})$ 是一个有界线性映射。附录 A 简要回顾了与向量值核相关的关键定义，另见Carmeli 等人（2005年，2008年）。

假设 2.1（核）。 假设核 K 具有如下形式的积分表示

$$K(u, \tilde{u}) = \sum_{i=1}^{p} \int_\Omega \varphi_i(u, \omega) \otimes \varphi_i(\tilde{u}, \omega)\, d\pi(\omega), \tag{2.5}$$

其中 $\varphi_i\colon \mathcal{U}\times\Omega \to \mathcal{V}$对于 $i = 1, \ldots, p$，且 (Ω, π) 是一个概率空间。此外，假设对于所有 $u \in \mathcal{U}$,

$$\sum_{i=1}^{p} \|\varphi_i(u, \omega)\|^2_\mathcal{V} \le \kappa^2 \quad \pi - almost\ surely\ . \tag{2.6}$$

注意，假设 2.1 蕴含希尔伯特–施密特界 $\|K(u, \tilde{u})\|_{HS} \le \kappa^2 \rho_\mathcal{U}$几乎必然成立。(2.5) 的例子包括高斯核和随机傅里叶特征（Rahimi and Recht, 2007年；Rudi 和 Rosasco，2016年）。与这些工作不同，我们允许对 i进行额外求和，从而使我们能够涵盖神经切线核 (2.2) 的特殊情形，另见 Jacot 等人（2018）；Nitanda 和 Suzuki（2020）；Li 等人（2021a）；Munteanu 等人（2022）；Oymak 和 Soltanolkotabi（2019）；Nguyen 和 M¨ucke（2023）。

随机特征近似。RFA 的基本思想是，将具有积分表示（2.5）的核近似为由有限项求和表示的核，即 $K(u, \tilde{u}) \approx K_M(u, \tilde{u})$，其中 $u, \tilde{u} \in \mathcal{U}$ 且 $M \in \mathbb{N}$，其中

$$K_M(u, \tilde{u}) := \sum_{i=1}^{p} \frac{1}{M} \sum_{m=1}^{M} \varphi_i(u, \omega_m) \otimes \varphi_i(\tilde{u}, \omega_m).$$

这里， $\varphi_i : \mathcal{U}\times\Omega \to \mathcal{V}$ 是某个概率空间（Ω, π），而 $\{\omega_m\}_{m=1}^M$ 是从 π中独立同分布抽取的。
相应的再生核希尔伯特空间记为 $\mathcal{H}_M$。在假设（2.6）下，嵌入映射 $S_M : \mathcal{H}_M \hookrightarrow L^2(\mathcal{U}, \rho_\mathcal{U}; \mathcal{V})$是一个有界线性映射。

RFA 的主要优势在于，它使核方法能够在大型数据集上进行计算，同时保留其统计保证。经典核方法需要存储完整的 $n \times n$格拉姆矩阵，并求解线性系统，内存开销为$O(n^2)$，时间开销为 $O(n^3)$。相比之下，RFA 用显式的M维特征映射替代核，将计算开销降至 $O(nM^2)$（用于岭回归）或$O(nMt)$（用于 t 次梯度下降迭代），内存开销仅为 $O(nM)$。从统计角度来看，研究已表明，对于许多核，只需 $M = O(\sqrt{n})$个特征即可达到极小极大学习率。因此，RFA 为核方法提供了一个可扩展且灵活的框架。

正则化。 由于无法直接最小化期望风险（2.1），标准做法是在假设空间 $\mathcal{H}$上进行经验风险最小化（ERM）

$$\min_{F\in\mathcal{H}} \widehat{\mathcal{E}}(F), \qquad \widehat{\mathcal{E}}(F) = \frac{1}{n}\sum_{j=1}^{n} \ell\big(F(u_j), v_j\big).$$

然而，在再生核希尔伯特空间中直接进行经验风险最小化通常是不适定的。为避免过拟合并获得一致估计量，引入正则化 。

定义 2.2（正则化函数）。令 $\phi{:}(0,1]\times[0,1]\to\mathbb{R}$ 并设 $\phi_\lambda(t)=\phi(\lambda,t)$。若存在常数 $D, E, c_0>0$，使得对所有 $0<\lambda\le 1$ 都有：则称族 $\{\phi_\lambda\}_\lambda$ 为一族正则化函数。

$$\textit{i)}\quad \sup_{0<t\le 1}|t\phi_\lambda(t)|\le D, \tag{2.7}$$

$$\textit{ii)}\quad \sup_{0<t\le 1}|\phi_\lambda(t)|\le \tfrac{E}{\lambda}, \tag{2.8}$$

$$\textit{iii)}\quad \sup_{0<t\le 1}|r_\lambda(t)|\le c_0,\ \ r_\lambda(t):=1-t\phi_\lambda(t). \tag{2.9}$$

这类方法统称为谱正则，既包括 Tikhonov 正则化等显式正则化，也包括通过迭代方案实现的隐式正则化，例如梯度下降和加速方法。这些技术最初是为（统计）逆问题而开发的（Engl 等人，1996年），此后被应用于机器学习，尤其是非参数最小二乘回归（Caponnetto 和 De Vito，2007年；Bauer 等人，2007年；Blanchard 和 M¨ucke，2017；Lin 等人，2020b年）。

Gerfo 等人在(2008年)；Blanchard 和 M¨ucke (2017) 的研究表明，可达到的学习率实质上由正则化的资格决定 $\{\phi_\lambda\}_\lambda$，即满足以下条件的最大 $\nu>0$，使得对所有 $q\in[0,\nu]$ 和 $0<\lambda\le 1$:

$$\sup_{0<t\le 1}|r_\lambda(t)|\,t^q\le c_q\,\lambda^q, \tag{2.10}$$

对于某个常数 $c_q>0$。

一种有原则的方法是利用经验算子 $\Sigma_M:\mathcal{H}_M\to\mathcal{H}_M$ 和 $\mathcal{S}_M^*:\widehat{\mathcal{V}^n}\to\mathcal{H}_M$ 的谱结构，其定义如下：

$$\widehat{\Sigma}_M=\frac{1}{n}\sum_{j=1}^n K_{M,u_j}K_{M,u_j}^*,$$

$$\widehat{\mathcal{S}}_M^*\mathbf{v}=\frac{1}{n}\sum_{j=1}^n K_{M,u_j}v_j.$$

利用这些算子，结合 RFA 的谱正则化估计量具有如下形式：

$$F_\lambda^M=\phi_\lambda(\widehat{\Sigma}_M)\,\widehat{\mathcal{S}}_M^*\mathbf{v}\ \in\ \mathcal{H}_M. \tag{2.11}$$

3 主要结果

3.1 假设与主要结果

本节将阐述我们的假设并陈述主要结果。

假设 3.1（数据分布）。存在正数常数 Q和 Z，使得对于所有满足 $l\in\mathbb{N}$的$l\ge 2$，

$$\int_{\mathcal{V}}\|v\|_{\mathcal{V}}^l\,d\rho(v\mid u)\le\frac{1}{2}l!Z^{l-2}Q^2$$

$\rho_{\mathcal{U}}$-几乎必然。

例如，若 v 几乎必然有界，则该假设成立。这进一步意味着回归算子 G_ρ 几乎必然有界，因为

$$\begin{aligned}\|G_\rho(u)\|_{\mathcal{V}}&\le\int_{\mathcal{V}}\|v\|_{\mathcal{V}}\,d\rho(v\mid u)\\&\le\left(\int_{\mathcal{V}}\|v\|_{\mathcal{V}}^2\,d\rho(v\mid u)\right)^{\frac{1}{2}}\le Q\,.\end{aligned}$$

为了刻画 G_ρ 相对于核的光滑性，我们施加一个所谓的源条件。该条件将 G_ρ 与核积分算子的谱性质联系起来，在确定可达到的学习率方面起着核心作用。记 $\mathcal{L}$为：$L^2(\mathcal{U},\rho_{\mathcal{U}})\to L^2(\mathcal{U},\rho_{\mathcal{U}})$，即与 K相关联的核积分算子，

$$\mathcal{L}G=\int_{\mathcal{U}}K_uG(u)\,\rho_{\mathcal{U}}(du).$$

假设 3.2（源条件）。设 $R>0,r>0$。我们假设 $G_\rho=\mathcal{L}^rH$，其中某个 $H\in L^2(\mathcal{U},\rho_{\mathcal{U}})$ 满足$\|H\|_{L^2(\rho_{\mathcal{U}})}\le R$。

该条件将 G_ρ 与核积分算子的谱性质联系起来，在确定可达到的学习率方面起着核心作用。参数 $r>0$ 量化了 G_ρ的光滑程度：r 值越大表示正则性越高。特别地，情形 $r=\frac{1}{2}$ 对应于设定正确的情形，其中 G_ρ属于 $\mathcal{H}$；而 $r>\frac{1}{2}$ 则体现了额外的光滑性，$r<\frac{1}{2}$ 对应于设定错误的情形。更多细节可参见例如Bauer 等人（2007年）；Lin 等人（2020b年）。

虽然源条件控制着目标函数的正则性，但仍需要一个互补假设来刻画假设空间的复杂度。这通常借助有效维数来表述，它衡量再生核希尔伯特空间相对于核算子谱的容量。

假设 3.3（有效维数）。对某些 $b\in[0,1]$ 和 $c_b>0$，假设对所有$\lambda>0$，算子 $\mathcal{L}$ 满足

$$\mathcal{N}(\lambda):=\operatorname{tr}\left(\mathcal{L}(\mathcal{L}+\lambda I)^{-1}\right)\le c_b\lambda^{-b}. \tag{3.1}$$

此外，我们假设 $2r+b>1$。

此处，$\mathcal{N}(\lambda)$ 是核的有效维数，它量化了假设空间的有效自由度数量。直观而言，它反映了 $\mathcal{L}$的特征值衰减情况：b值越小，衰减越快（容量越低）；而 b 值越大，则表示衰减越慢，因此复杂度越高。条件（3.1）在 $b=1$时总是成立，因为 $\mathcal{L}$ 是迹类，且其特征值 $\{\mu_i\}$ 满足 $\mu_i\lesssim i^{-1}$。更一般地，若特征值按多项式规律衰减：

$\mu_i \sim i^{-c}$ 且 $c > 1$，则（3.1）在 $b = 1/c$时成立；若 $\mathcal{L}$ 的秩有限，则 $b = 0$。情形 $b = 1$ 通常称为与容量无关的情形。较小的 b 值有助于学习算法实现更快的收敛率。

正则化参数 λ以及随机特征的数量 M 在（2.11）中至关重要，用于平衡近似误差、估计误差和优化误差（见附录 B）。在实践中，这两个参数都根据样本量 n确定，以保证最优的统计性能。以下结果表明在满足源条件和容量条件的情况下，我们的随机特征估计量达到了极小极大最优收敛速率。此外，它还给出了明确的关于 λ 的条件 n 以及 M_n 确保所需的统计保证。证明见附录 B。

定理 3.4。假设 *3.1–3.3*成立。设 $\{\phi_\lambda\}_\lambda$ 是一族正则化函数，其资格为 $\nu > 0$。令 $\delta \in (0,1)$ 并选择

$$\lambda_n = C\, n^{-\frac{1}{2r+b}} \log^3\!\left(\tfrac{2}{\delta}\right).$$

那么，以至少为 $1-\delta$ 的概率，随机特征估计量(2.11)满足

$$\|G_\rho - \mathcal{S}_{M_n} F_{\lambda_n}^{M_n}\|_{L^2(\rho_\mathcal{U})} \le \bar{C}\, n^{-\frac{r}{2r+b}} \log^{3r+1}\!\left(\tfrac{1}{\delta}\right),$$

前提是 $\nu \ge r \vee 1$、$n \ge n_0 := \exp\binom{2r+b}{2r+b-1}$，且随机特征的数量满足

$$M_n \ge p \cdot \tilde{C} \cdot \log(n) \cdot \begin{cases} n^{\frac{1}{2r+b}}, & r \in (0, \frac{1}{2}), \\ n^{\frac{1+b(2r-1)}{2r+b}}, & r \in [\frac{1}{2}, 1], \\ n^{\frac{2r}{2r+b}}, & r \in (1, \infty). \end{cases}$$

这里 $\tilde{C}$、$\bar{C}$、C 是与 n、M、λ 无关的常数。

由于我们的框架涵盖了通过梯度下降训练的算子值随机特征近似，定理 3.4 可直接用于推导神经算子的泛化界。回顾浅层神经算子在（2.3）中的超额风险分解。Nguyen 和 M¨ucke(2024)表明第一项的上界为$O(\log n/M_n)$。因此，当神经元数量随随机特征数量缩放，即 $M_n = O\left(n^{\frac{2r}{2r+b}} \log n\right)$ 时，定理 3.4 表明神经算子达到与非参数核方法相同的极小极大速率。

推论 3.5 (Nguyen 和 M¨ucke (2024)，定理 3.5)。假设定理 *3.4* 中的假设成立。令 $G_{\theta_{T_n}}$ 表示第 *2.1*节中定义的 *NO*，其中 T_n 表示梯度下降迭代次数。令 M_n 表示网络宽度。假设

$$\lambda_n = T_n^{-1} = C\, n^{-\frac{1}{2r+b}},$$
$$M_n \ge \tilde{d}^2 \cdot \tilde{C}\, B_{T_n}^6 \left(T_n^{2r} \vee T_n\right) \log^2 n,$$

其中 $B_{T_n} > 0$ 界定了参数漂移，

$$\|\theta_t - \theta_0\|_2 \le B_{T_n} \quad for\ all\ t \in [T_n],$$

与 C、$\bar{C}$ 是与n、M_n、T_n、B_{T_n}无关的正数常数。于是，以至少$1-\delta$的概率，

$$\|G_{\theta_{T_n}} - G_\rho\|_{L^2(\rho_\mathcal{U})} \le \bar{C}\, n^{-\frac{r}{2r+b}} \log^3(2/\delta),$$

，其中常数 $C > 0$ 与n、M_n、T_n、B_{T_n}无关。

关于训练、θ_0的初始化以及上述推论的更多细节见附录A。

3.2 讨论

定理3.4。定理3.4 针对一大类谱滤波方法建立了极小极大速率。在良设定情形（$r = \frac{1}{2}$，$b = 1$）下，若要达到平方 L^2误差界，其阶为 $O(1/\sqrt{n})$，则所需迭代次数为 $t_n = 1/\lambda_n = O(\sqrt{n})$，随机特征数为 $M_n = O(\sqrt{n}\log n)$。对于正则性为 $r \ge 1$的更光滑目标函数，最优收敛速率以 $t_n = O(n^{\frac{1}{2r+1}})$ 的迭代次数达到，但需要 $M_n = O(t_n^{2r} \log n)$ 个随机特征。这揭示了一个有趣的权衡：更高的光滑性会减少必要的迭代次数，但会增加实现最优泛化所需的随机特征数量。

相比之下，在设定错误情形下 $r < \frac{1}{2}$，可达到的速率为 $O(n^{-\frac{r}{2r+1}})$，速度较慢，反映了目标函数正则性的有限性。在这一情形下，所需的随机特征数量仅为 $M_n = O(n^{\frac{1}{2r+1}} \log n)$，显著少于设定正确或光滑情形。总的来说，定理3.4 表明，采用谱正则化的随机特征方法能够达到与精确核方法相同的极小极大最优速率（Caponnetto 和 De Vito，2007年；Blanchard 和 M¨ucke，2017），同时具有更好的计算可扩展性。

与先前工作的比较。与Rudi 和 Rosasco(2016年)；Lanthaler 和 Nelsen (2023的结果相比，我们将分析从核岭回归扩展到一般的谱滤波方法，并针对所有满足 $r < \frac{1}{2}$ 的光滑度级别 $2r + b > 1$ （容易学习情形）建立了最优收敛率。就所需随机特征的数量而言，我们得到了与Rudi 和 Rosasco(2016年)相同的阶，即$M_n = O(\sqrt{n}\log n)$。Lanthaler 和 Nelsen(2023)的分析基于略有不同的源条件，在设定正确的情形下，这些条件与我们的条件一致。他们采用随机厨房水槽方法（Rahimi andRecht，2008年），进一步证明可以去掉对数因子，从而证明只需 $M_n = O(\sqrt{n}$个随机特征即可

达到最优收敛率。然而，他们的结果没有利用关于有效维数的先验知识，因此只在设定正确的情形下证明了最优收敛率 $b = 1, r = \frac{1}{2}$。

神经网络与神经算子。 神经网络学习与随机特征近似（RFA）之间的联系，已由 Yehudai 和 Shamir (2019) 观察到。粗略地说，他们指出，只要随机特征学习可行，神经网络学习也可行。同时，他们证明，在高斯输入下，若网络规模（或权重幅度）不是关于 $\mathbb{R}^d$ 的多项式级别，则神经网络无法学习哪怕单个 ReLU神经元，除非网络规模（或权重幅度）相对于 d 呈指数级增长。在神经切线核区域内使用更平滑的激活函数时，Nguyen 和 Md¨ucke (2023) 对这一结果作了改进，证明只需多项式数量的随机特征，就能在输入维度 d 和样本复杂度 n方面达到最优。

我们的方法将这些见解扩展到了与神经算子（NOs）相关的算子值设定。与向量输入情形不同，我们的收敛率不依赖于输入空间的维数 $\mathcal{U}$。然而，核表示的求和结构（2.5）使其对加项数量呈线性依赖 p，如定理3.4所示。对于神经算子，输入空间 $\mathcal{U}$ 通常是一个函数空间，例如从$\mathcal{X}$ 到 $\mathbb{R}^{d_y}$的连续映射。在这种情况下，输入函数的输出维度 d_y 会进入特征维˜数 $d = d_k + d_y + d_b$，并直接决定所需随机特征的数量。因此，我们的界整体上呈现出 $\tilde{d}^2$ 量级的依赖（见第 2.1节），如推论 3.5 所述。因此，我们的结果揭示了明显的权衡：泛化率不依赖于 $\mathcal{U}$（可能是无限的）维数，但计算成本˜随特征维数呈二次增长 d。

4 结论

我们针对向量值核的 RFA 构建了统一的谱滤波框架，旨在为神经算子推导泛化保证。我们的分析不局限于核岭回归，还涵盖了具有显式或隐式正则化的广泛学习算法，并将以往结果作为特例纳入其中。我们方法的一项关键优势是，收敛率和特征需求均不依赖于（可能是无限维的）输入空间维数，因此结果可直接应用于神经算子。同时，我们的界限仅随每个神经元的特征维数呈二次增长，从而首次为神经算子提供了兼顾统计效率与计算可处理性的极小极大最优保证。

主要理论结果——定理3.4，确立了在标准的源条件和容量条件下，RFA 达到了极小极大速率，与精确核方法相匹配，同时所需资源显著更少。我们的讨论重点阐述了光滑性、迭代复杂度与随机特征数量之间的权衡，以及良好设定、光滑和设定错误情形之间的差异。与先前工作相比，我们的框架在容易学习情形下为所有 $r < \frac{1}{2}$ 给出了最优速率，并涵盖了算子值神经切线核，在统一的理论框架中连接了神经网络、随机特征和神经算子。未来一个有趣的研究方向是探究能否进一步降低特征维数 d 的二次依赖˜，并将分析从神经切线核区域扩展到更深层架构。

参考文献

A. B. Aleksandrov 和 V. V. Peller. 算子 Hölder–Zygmund 函数，2009年。网址 ¨older–zygmund 函数，2009年。网址 https://arxiv.org/abs/0907.3049。 B.6Frank Bauer、Sergei Pereverzev 和 Lorenzo Rosasco. 学习理论中的正则化算法。 复杂性杂志，23(1):52–72，2007年。 2.2，3.1Gilles Blanchard 和 Nicole M¨ucke. 统计逆学习问题正则化的最优收敛率。 计算数学基础，18:971–1013，2017年。 2.2，3.2，B.1，B.6Romain Brault、Florence d'Alch´e Buc 和 Markus Heinonen. 面向算子值核的随机傅里叶特征，2016年。网址 https://arxiv.org/abs/1605.02536。 A.1A. 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: 331–368，2007年。 1，2.2，3.2，B.20

C. Carmeli、E. De Vito 和 V. Umanit`a A. Toigo。向量值再生核希尔伯特空间及其普适性，2008年。2，2.2，A.1

Claudio Carmeli、Ernesto De Vito 和 Alessandro Toigo。再生核希尔伯特空间与 Mercer 定理，2005年。2，2.2，A.1Luigi Carratino、Alessandro Rudi 和 Lorenzo Rosasco。使用随机梯度下降和随机特征进行学习，2019年。1Heinz Werner Engl、Martin Hanke 和 Andreas Neubauer。逆问题的正则化，第 375 卷。Springer Science & Business Media，1996年。2.2L. Lo Gerfo、L. Rosasco、F. Odone、E. De Vito 和 A. Verri。用于监督学习的谱算法。神经计算，20(7):1873–1897，2008年。ISSN 0899-7667。doi: 10.1162/neco.2008.

05-07-517。网址 https://doi.org/10.1162/neco.2008.05-07-517。2.2 Daniel Zhengyu Huang、Nicholas H. Nelsen 和 Margaret Trautner。参数到可观测量映射的算子学习视角，2024。1 Arthur Jacot、Cl´ement Hongler 和 Franck Gabriel。神经切线核：神经网络中的收敛与泛化。载于 NeurIPS，2018。2.1、2.2 Taeyoung Kim 和 Myungjoo Kang。傅里叶神经算子的 Rademacher 复杂度界，2022。网址 https://arxiv.org/abs/ 2209.05150。1 Nikola Kovachki、Samuel Lanthaler 和 Siddhartha Mishra。关于傅里叶神经算子的万能逼近与误差界。机器学习研究期刊，22(290):1–76，2021。网址 http://jmlr.org/papers/ v22/21-0806.html。1 Nikola Kovachki、Zongyi Li、Burigede Liu、Kamyar Azizzadenesheli、Kaushik Bhattacharya、Andrew Stuart 和 Anima Anandkumar。神经算子：学习函数空间之间的映射及其在偏微分方程中的应用。机器学习研究期刊，24(89):1–97，2023a。1 Nikola B. Kovachki、Zong-Yi Li、Burigede Liu、Kamyar Azizzadenesheli、Kaushik Bhattacharya、Andrew M. Stuart 和 Anima Anandkumar。神经算子：学习函数空间之间的映射及其在偏微分方程中的应用。J. Mach. Learn. Res.，24:89:1–89:97，2023b。网址 https://api.semanticscholar.org/CorpusID:259149906。1 Nikola B. Kovachki、Samuel Lanthaler 和 Andrew M. Stuart。算子学习：算法与分析，2024。1 Samuel Lanthaler 和 Nicholas H. Nelsen。使用向量值随机特征进行学习的误差界，2023。1、1、3.2、A.1 Jose Antonio Lara Benitez、Takashi Furuya、Florian Faucher、Anastasis Kratsios、Xavier Tricoche 和 Maarten V. de Hoop。神经算子的分布外风险界及其在亥姆霍兹方程中的应用。《计算物理学杂志》，513:113168，2024。ISSN 0021-9991。doi: https://doi.org/10.1016/j.jcp.2024.113168。网址 https://www.sciencedirect.com/science/article/pii/S0021999124004170。1 Jaehoon Lee、Lechao Xiao、Samuel Schoenholz、Yasaman Bahri、Roman Novak、Jascha Sohl-Dickstein 和 Jeffrey Pennington。任意深度的宽神经网络在梯度下降下演化为线性模型。《神经信息处理系统进展》，32，2019。2.1 Mufan Bill Li、Mihai Nica 和 Daniel M. Roy。未来是对数高斯：残差网络及其在初始化时的无限深度与宽度极限，2021a。2.2 Zhu Li、Jean-Francois Ton、Dino Oglic 和 Dino Sejdinovic。迈向随机傅里叶特征的统一分析，2021b。1 Zongyi Li、Nikola Kovachki、Kamyar Azizzade- nesheli、Burigede Liu、Kaushik Bhattacharya、Andrew Stuart 和 Anima Anandkumar。神经算子：用于偏微分方程的图核网络，2020。网址：https: // arxiv.org/abs/2003.03485。1 Zongyi Li、Nikola Borislavov Kovachki、Kamyar Azizzadenesheli、Burigede liu、Kaushik Bhat- tacharya、Andrew Stuart 和 Anima Anand- kumar。用于参数化偏微分方程的傅里叶神经算子。发表于国际学习表征会议，2021c。网址：https://openreview.net/forum? id= c8P9NQVtmnO。1 Junhong Lin 和 Volkan Cevher。采用随机梯度方法和谱算法的分布式学习的最优收敛性，2018。网址：https://arxiv.org/abs/1801.07226。1，B.10，B.6，B.6 Junhong Lin、Alessandro Rudi、Lorenzo Rosasco 和 Volkan Cevher。希尔伯特空间上最小二乘回归的谱算法最优速率。《应用与计算调和分析》，48(3):868–890，2020a。1 Junhong Lin、Alessandro Rudi、Lorenzo Rosasco 和 Volkan Cevher。希尔伯特空间上最小二乘回归的谱算法最优速率。《应用与计算调和分析》，48(3):868–890，2020b。2.2，3.1 Carlo Marcati 和 Christoph Schwab。椭圆偏微分方程深度算子网络的指数收敛。《SIAM 数值分析杂志》，61(3):1513–1545，2023年6月。ISSN 1095-7170。doi：10.1137/21m1465718。网址：http://dx.doi.org/10.1137/21M1465718。1 Ha Quang Minh。算子值 Bochner 定理、算子值核的傅里叶特征映射及向量值学习，2016。网址：https://arxiv.org/abs/1608.05639。A.1 Mattes Mollenhauer、Nicole M¨ucke 和 T. J. Sullivan。学习线性算子：无限维回归作为适定的非紧逆问题，2024。网址：https://arxiv.org/abs/2211.08875。A.1 Alexander Munteanu、Simon Omlor、Zhao Song 和 David P. Woodruff。通过耦合初始化界定神经网络的宽度——最坏情形分析，2022。2.2

Mike Nguyen 和 Nicole M¨ucke。我们需要多少个神经元？对使用梯度下降训练的浅层网络的精细分析，2023。2.2，3.2Mike Nguyen 和 Nicole M¨ucke。神经算子的最优收敛速率，2024。网址 https://arxiv.org/abs/2412.17518。1，2.1，2.1，3.1，3.5，A.2，A.2，A.2，A.1，A.2Atsushi Nitanda 和 Taiji Suzuki。神经切线核机制下平均随机梯度下降的最优速率。发表于国际学习表征会议。arXiv，2020。2.2

Samet Oymak 与 Mahdi Soltanolkotabi。迈向适度过参数化：训练浅层神经网络的全局收敛保证，2019。2.2

Nicol`o Pagliana 和 Lorenzo Rosasco。《希尔伯特空间中加速方法的隐式正则化》。神经信息处理系统进展，32:14481–14491，2019。1 Shaoxiang Qin、Fuyuan Lyu、Wenhui Peng、Dingyang Geng、Ju Wang、Xing Tang、Sylvie Leroyer、Naiping Gao、Xue Liu 和 Liangzhu Leon Wang。《从谱的角度更好地理解傅里叶神经算子》，2024。网址：https://arxiv.org/abs/2404.07200。1 Ali Rahimi 和 Benjamin Recht。《面向大规模核机器的随机特征》。载于《神经信息处理系统进展》。Curran Associates, Inc.，2007。1, 1, 2.2 Ali Rahimi 和 Benjamin Recht。《随机厨房水槽的加权和：以随机化取代学习中的最小化》。载于 D. Koller、D. Schuurmans、Y. Bengio 和 L. Bottou 编，《神经信息处理系统进展》，第21卷。Curran Associates, Inc.，2008。网址：https://proceedings. neurips.cc/paper_files/paper/2008/file/0efe32849d230d7f53049ddc4a4b0c60-Paper.pdf。3.2 M. Raissi、P. Perdikaris 和 G.E. Karniadakis。《物理信息神经网络：用于求解涉及非线性偏微分方程的正问题和逆问题的深度学习框架》。《计算物理学杂志》，378:686–707，2019。ISSN 0021-9991。doi：https://doi.org/10.1016/j.jcp.2018.10.045。网址：https://www.sciencedirect.com/science/article/pii/S0021999118307125。1 Alessandro Rudi 和 Lorenzo Rosasco。《随机特征学习的泛化性质》，2016。1, 1, 2.2, 3.2, B.1, B.8, B.5, B.13, B.6 B. Schoelkopf 和 A. J. Smola。《核学习：支持向量机、正则化、优化及其他（自适应计算与机器学习）。麻省理工学院出版社，2002。1

Christoph Schwab 与 Jakob Zech。高维深度学习：解析函数的神经网络逼近，见于 $l^2(\mathbb{R}^d, \gamma_d)$，2021。网址https://arxiv.org/abs/2111.07080。1

Jun Shao。数理统计。Springer-Verlag New York Inc，第2版，2003。2
Anil Sharma、Suresh Singh 与 Sameer Ratna。图神经网络算子：综述。多媒体工具与应用，83:23413–23436，2024。doi: 10.1007/s11042-023-16440-4。1

Ingo Steinwart 和 Andreas Christmann。支持向量机。Springer Science & Business Media，2008年。A.1，B.4Joel A. Tropp。随机矩阵之和的易用尾界。计算数学基础，12(4):389–434，2011年。B.6Sifan Wang、Hanwen Wang 和 Paris Perdikaris。利用物理信息深度算子网络学习参数化偏微分方程的解算子。《科学进展》，7(40): eabi8605，2021年。doi: 10.1126/sciadv.abi8605。网址：
h
t
t
p
s://
w
ww.science.org/doi/abs/10.1126/sciadv.abi8605。1Gilad Yehudai 和 Ohad Shamir。论随机特征在理解神经网络方面的能力与局限。神经信息处理系统进展，32，2019年。3.2Haobo Zhang、Yicheng Li 和 Qian Lin。论错误设定谱算法的最优性，2024。网址：https://arxiv.org/abs/2303.14942。1Yaoyu Zhang、Zhi-Qin John Xu、Tao Luo 和 Zheng Ma。深度神经网络中由初始化引发的一类泛化误差。载于数学与科学机器学习，第 144–164 页。PMLR，2020年。A.2Xiantong Zhen、Haoliang Sun、Yingjun Du、Jun Xu、Yilong Yin、Ling Shao 和 Cees Snoek。利用变分随机特征学习核，2020年。1

A 补充材料

本附录首先回顾与向量值核相关的关键定义。随后，我们补充介绍神经算子，并给出推论3.5的更详细版本，同时简要讨论其含义。最后，我们提供数值示例，以支持主要理论结果——定理3.4。

A.1 向量值核预备知识

实值再生核希尔伯特空间的经典理论（Steinwart 和 Christmann，2008年），包括 Mercer 定理等基本结果，已在 Carmeli 等人 （2005，2008）中扩展到向量值情形。这些扩展为在核框架内分析算子学习问题提供了严格的数学基础（Minh，2016年；Brault 等人，2016年；Mollenhauer 等人，2024；Lanthaler 和 Nelsen，2023），并已成为算子学习文献中的标准工具。为完整起见，下面简要回顾关键定义。

设 $\mathcal{U}$ 为一个拓扑空间，设 $\mathcal{V}$ 为一个可分希尔伯特空间。映射

$$K : \mathcal{U} \times \mathcal{U} \to \mathcal{L}(\mathcal{V}),$$

其中， $\mathcal{L}(\mathcal{V})$ 表示 $\mathcal{V}$上的有界线性算子空间，若对于任意有限点集 u_1，……， $u_N \in \mathcal{U}$以及向量 v_1，……，$v_N \in \mathcal{V}$，均有下式成立，则称其为 $\mathcal{V}$-再生核在 $\mathcal{U}$ 上的

$$\sum_{i,j=1}^{N} \langle K(u_i, u_j)v_j, v_i\rangle_{\mathcal{V}} \geq 0.$$

这一条件是正定性从标量值情形推广到向量值情形的自然推广。

对于每个 $u \in \mathcal{U}$，我们定义线性算子

$$K_u : \mathcal{V} \to \mathcal{F}(\mathcal{U}; \mathcal{V}),$$

其中 $\mathcal{F}(\mathcal{U}; \mathcal{V})$ 表示 $\mathcal{U}$ 上取值于 $\mathcal{V}$ 的可测函数空间。其在 $v \in \mathcal{V}$ 上的作用定义为

$$(K_u v)(\tilde{u}) = K(\tilde{u}, u)v \qquad \text{for all } \tilde{u} \in \mathcal{U}.$$

给定一个 $\mathcal{V}$-再生核 K，存在唯一的希尔伯特空间 $\mathcal{H}_K \subset \mathcal{F}(\mathcal{U}; \mathcal{V})$，使得

$$K_u \in \mathcal{L}(\mathcal{V}, \mathcal{H}_K) \quad \text{for all } u \in \mathcal{U},$$

并且对于每个 $F \in \mathcal{H}_K$，

$$F(u) = K_u^* F \qquad \text{for all } u \in \mathcal{U},$$

其中 $K_u^* : \mathcal{H}_K \to \mathcal{V}$ 表示 K_u 的伴随算子。这是经典再生性质的向量值类比。特别地，它意味着

$$K(u, \tilde{u}) = K_u^* K_{\tilde{u}}.$$

空间 $\mathcal{H}_K$ 称为与 K相关联的向量值再生核希尔伯特空间（*RKHS*），其定义为

$$\mathcal{H}_K = \overline{\text{span}}\{K_u v \,|\, u \in \mathcal{U},\, v \in \mathcal{V}\}.$$

最后，再生核 $K\colon \mathcal{U} \times \mathcal{U} \to \mathcal{L}(\mathcal{V})$称为*Mercer*核，如果 $\mathcal{H}_K$ 是 $\mathcal{C}(\mathcal{U}; \mathcal{V})$的子空间，即定义在 $\mathcal{U}$上的连续 $\mathcal{V}$值函数空间。

A.2 使用神经算子学习

本节将详细阐述推论 3.5，该推论基于 Nguyen 和 M¨ucke (2024，定理 3.5）。为求完整起见，我们简要回顾他们工作中所考虑的神经算子学习设定，包括通过梯度下降进行训练的过程。

神经算子设定。 我们考虑第 2 节介绍的一般算子学习框架，其中输入空间和输出空间 $\mathcal{U}$ 和 $\mathcal{V}$ 均被视为无限维函数空间。为了基于函数值数据进行实际训练，还需要额外的离散化步骤。这涉及从函数的输入域中抽取所谓的第二阶段样本 。

具体而言，令 $\mathcal{X} \subseteq \mathrm{R}^{d_x}$ 表示输入域，并令 μ 为定义在 $\mathcal{X}$上的未知概率测度。空间 $\mathcal{U}$ 和 $\mathcal{V}$ 分别由从 $\mathcal{X} \to \mathcal{Y} \subset$ 映射到 R^{d_y} 、从 $\mathcal{X} \to \tilde{\mathcal{Y}} \subset$映射到 R 的函数构成。我们观测到 n_u 个独立同分布的第一阶段样本

$$\{(u_i, v_i)\}_{i=1}^{n_u} \subset \mathcal{U} \times \mathcal{V},$$

每个样本都在 n_x 个独立同分布的第二阶段样本 (x_1、……、x_{n_x}) $\in \mathcal{X}^{n_x}$ 上进行评估。随后使用这些离散化的评估结果，通过梯度下降训练浅层神经算子。

梯度下降与初始化。 回顾第 2.1 节介绍的浅层神经算子类。根据 Nguyen 和 M¨ucke (2024)，通过梯度下降，利用第一阶段样本在第二阶段点处的评估值计算经验损失，并据此训练网络参数：

$$\begin{aligned} \theta_{t+1}^j &= \theta_t^j - \alpha\, \partial_{\theta^j} \widehat{\mathcal{E}}\big(G_{\theta_t}\big) \\ &= \theta_t^j - \frac{\alpha}{n_u} \sum_{i=1}^{n_u} \Big\langle G_{\theta_t}(u_i) - v_i,\, \partial_{\theta^j} G_{\theta_t}(u_i) \Big\rangle_{n_x}, \end{aligned} \tag{A.1}$$

其中 $\alpha > 0$ 为步长，并且

$$\langle f, g \rangle_{n_x} := \frac{1}{n_x} \sum_{k=1}^{n_x} f(x_k) g(x_k)$$

表示第二阶段样本上的经验内积。

Nguyen 和 M¨ucke (2024) 对网络参数采用对称初始化方案 θ_0 以确保 $G_{\theta_0} \equiv 0$。重要的是，这种对称技巧不会影响极限神经切线核（NTK）；详情请参阅 Zhang等人 (2020)。

具体而言，输出层中的权重按对称方式初始化为

$$a_m^{(0)} = \tau \quad \text{for } m = 1, \ldots, M/2, \qquad a_m^{(0)} = -\tau \quad \text{for } m = M/2+1, \ldots, M,$$

其中 $\tau > 0$ 是一个固定常数。输入层参数以耦合方式初始化，

$$b_m^{(0)} = b_{m+M/2}^{(0)} \quad \text{for } m \in \{1, \ldots, M/2\},$$

其中，参数 $\{b_m^{(0)}\}_{m=1}^{M/2}$ 的前一半独立地从初始化分布 π_0中抽取。

现在我们可以陈述Nguyen 和 M¨ucke （2024）提出的原始定理，该定理给出了通过经验梯度下降算法训练的神经算子的泛化界。

定理 A.1 （Nguyen 和 M¨ucke（2024），定理 3.5）。 假设*3.1*、*3.2*和*3.3* 均成立。设 $G_{\theta_{T_n}}$ 表示第*2.1* 节中定义的神经算子，其中 T_n 表示（A.1）中的梯度下降迭代次数。设M_{n_u}表示网络宽度。假设 $\alpha \in$（0，κ^{-2}），$n_u \geq n_0 := e^{\frac{2r+b}{2r+b-1}}$，并且

$$T_{n_u} = C n_u^{\frac{1}{2r+b}}, \qquad M_{n_u} \geq \tilde{C} B_{T_{n_u}}^6 \log^2(n_u) T_{n_u}^{2r \vee 1}, \qquad n_x \geq \tilde{C} B_{T_{n_u}}^2 T_{n_u}^{2r} \log^2 T_{n_u},$$

with

$$\|\theta_t - \theta_0\|_\Theta \leq B_{T_{n_u}} \quad \textit{for all } t \in [T_{n_u}]. \tag{A.2}$$

那么，以至少 $1 - \delta$ 的概率，

$$\|G_{\theta_{T_{n_u}}} - G_\rho\|_{L^2(\rho_\mathcal{U})} \leq \bar{C}\, n_u^{-\frac{r}{2r+b}} \log^3 \frac{2}{\delta}, \tag{A.3}$$

~ ¯其中 C、C、$C > 0$ 均独立于 n_u、n_x、M_{n_u}、T_{n_u}、$B_{T_{n_u}}$。

讨论。 Nguyen 和 M¨ucke (2024，定理 3.7）证明，对于$B_T = O(\log T)$，式（A.2）以高概率成立。因此，通过梯度下降训练的两层神经算子可达到¯极小极大最优学习率 $n_u^{-\frac{r}{2r+b}}$。值得注意的是，所需的网络宽度 M 与核梯度下降所需的随机特征数量相匹配，且与 n_x无关。 n_x 下界2.3）源于对误差分解中第一项的控制，该项刻画了

函数值样本的离散化。为确保式（A.1）中的经验梯度下降紧密跟踪总体梯度流，第二阶段样本上的经验内积必须一致逼近 $L^2(\mathcal{X},\ \mu)$ 内积。Hoeffding 型浓缩不等式给出一个 $O(n_x^{-1/2})$ 离散化误差，而该误差必须与极小极大最优学习率相匹配。这一要求导出了所述的 n_x 下界。

A.3 数值示例

我们分析实值神经正切核下核梯度下降的行为。在模拟中，我们使用从两个数据集中抽取的 $n = 5000$ 个训练样本和测试样本：（i）输入维度为 $d = 1$ 的标准正态分布；（ii）输入维度为 $d = 14$ 的 SUSY[1] 分类数据集的一个子集。所有报告的结果均为该算法 50 次独立运行的平均值。

我们的理论分析表明，数量级为 $M = O(\sqrt{n}\,p)$ 的随机特征就足以实现最优学习性能，其中$\bar{p} = d + 2$。图 1 显示，对于两个数据集，一旦 M 超过数量级为 $_O(\sqrt{n}\,p$ 的阈值，且梯度下降迭代次数 T 固定，进一步增加 M 不会使测试误差得到任何改善。

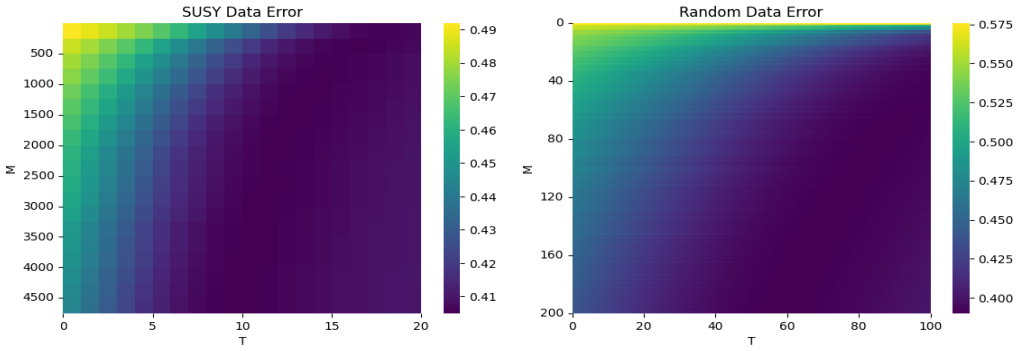

图 1：不同随机特征数量 M 和迭代次数 T 下的测试误差热图。**左图：** SUSY 数据集的误差。**右图：** 随机数据集的误差。

B 证明

本节给出主要结果的证明。

记号。证明过程中，我们使用以下简写记号。对于任意在两个希尔伯特空间 A 和 $\lambda > 0$之间映射的有界线性算子，我们记作 $A_\lambda := A + \lambda I$，其中 I 表示恒等算子。对于 $G\colon \mathcal{U} \to \mathcal{V}$和 $(u_1, ..., u_n) \in \mathcal{U}^n$，我们定义向量 $\bar{G} := \big(G(u_1) \ldots, G(u_n)\big) \in \mathcal{V}^n$。此外，令 $C_\bullet > 0$ 表示一个通用常数，它可能随行变化，但仅依赖于量 $\kappa, r, b, c_q, c_b, E, D, Q, Z$，而不依赖于 δ, p, λ或 n。

我们回顾以下算子定义。令 $\mathcal{S}_M\colon \mathcal{H}_M \hookrightarrow L^2(\mathcal{U}, \rho_\mathcal{U})$ 表示将$\mathcal{H}_M$嵌入到 $L^2(\mathcal{U}, \rho_\mathcal{U})$中的嵌入算子，其中 $M \in \mathbb{N}$。其伴随算子 $\mathcal{S}_M^*\colon L^2(\mathcal{U}, \rho_\mathcal{U}) \to \mathcal{H}_M$定义为

$$\mathcal{S}_M^* G = \int_\mathcal{U} K_{M,u}\, G(u)\, \rho_\mathcal{U}(\mathrm{d}u).$$

协方差算子 $\Sigma_M\colon \mathcal{H}_M \to \mathcal{H}_M$和核积分算子 $\mathcal{L}_M\colon L^2(\mathcal{U}, \rho_\mathcal{U}) \to L^2(\mathcal{U}, \rho_\mathcal{U})$ 定义为

$$\Sigma_M G := \int_\mathcal{U} K_{M,u} K_{M,u}^* G\, \rho_\mathcal{U}(\mathrm{d}u),$$
$$\mathcal{L}_M G := \int_\mathcal{U} K_{M,u} G(u)\, \rho_\mathcal{U}(\mathrm{d}u).$$

经验 这些算子的校准对应形式，通过将 $\rho_\mathcal{U}$ 替换为经验测度得到， re

[1] https://archive.ics.uci.edu/ml/datasets/SUSY

定义为

$$\begin{aligned}
\widehat{\mathcal{S}}_M : \mathcal{H}_M \to \mathcal{V}^n, \qquad & (\widehat{\mathcal{S}}_M G)_j = K^*_{M,u_j} G, \\
\widehat{\mathcal{S}}^*_M : \mathcal{V}^n \to \mathcal{H}_M, \qquad & \widehat{\mathcal{S}}^*_M \mathbf{v} = \frac{1}{n}\sum_{j=1}^{n} K_{M,u_j} v_j, \\
\widehat{\Sigma}_M : \mathcal{H}_M \to \mathcal{H}_M, \qquad & \widehat{\Sigma}_M = \frac{1}{n}\sum_{j=1}^{n} K_{M,u_j} K^*_{M,u_j}.
\end{aligned}$$

B.1 证明组织与误差分解

在不使用随机特征的经典核方法设定中，标准分析会引入理想化的总体估计量

$$F^*_\lambda := \mathcal{S}^*\, \phi_\lambda(\mathcal{L})\, G_\rho,$$

由此可将误差分解为偏差和方差两部分：

$$\begin{aligned}
G_\rho - \mathcal{S}\widehat{F}_\lambda &= \big(G_\rho - \mathcal{S}F^*_\lambda\big) + \big(\mathcal{S}F^*_\lambda - \mathcal{S}\widehat{F}_\lambda\big) \\
&= r_\lambda(\mathcal{L})\, \mathcal{L}^r H + \big(\mathcal{S}F^*_\lambda - \mathcal{S}\widehat{F}_\lambda\big),
\end{aligned}$$

其中，偏差通过残差多项式 r_λ进行控制，而方差则通过 Hoeffding 型集中不等式处理（Blanchard 和 M¨ucke，2017）。

在随机特征设定中，我们通过引入理想化估计量来调整这一方法

$$F^*_\lambda := \mathcal{S}^*_M\, \phi_\lambda(\mathcal{L}_M)\, G_\rho,$$

这自然会因有限维随机特征空间而引入额外的逼近误差 。具体而言，我们将超额风险分解为

$$\begin{aligned}
\|G_\rho - \mathcal{S}_M F^M_\lambda\|_{L^2(\rho_\mathcal{U})} &\le \|G_\rho - \mathcal{S}_M F^*_\lambda\|_{L^2(\rho_\mathcal{U})} + \|\mathcal{S}_M F^*_\lambda - \mathcal{S}_M F^M_\lambda\|_{L^2(\rho_\mathcal{U})} \\
&=: \text{Approximation Error} + \text{Estimation Error}. \qquad \text{(B.1)}
\end{aligned}$$

一个关键的技术难点在于控制逼近误差，这需要比较总体算子 $\mathcal{L}^r$ 与其随机特征对应算子 $\mathcal{L}^r_M$。以往的工作（Rudi 和 Rosasco，2016年）主要关注情形 $r \in [0.5, 1]$，而我们在第 B.5 节提出了新的算子不等式，从而能够处理任意 $r > 0$。对于估计项，经典分析依赖于核岭回归的特定结构。相比之下，我们的方法采用了精细的分解，利用残差 r_λ的多项式结构，从而能够对所有正则化滤波器一致地获得方差界。

我们分别在第 B.2 节和第 B.3节中界定逼近误差和估计误差，然后在第 B.4 节中结合这些界，证明主要结果定理 3.4。所需的算子不等式见第 B.5节，而必要的集中不等式汇总于第 B.6节。

B.2 界定逼近误差

命题 B.1. 假设 *3.1*、*2.1*、*3.2*、*3.3* 以及 $\nu \ge r \vee 1$ 成立。对于任意 $\lambda \in (0, 1]$,假设

$$M \;\ge\; p\, C_\bullet \log^2\!\big(\delta^{-1}\big)\, \log\!\big(\lambda^{-1}\big) \cdot \begin{cases} \lambda^{-1}, & r \in \big(0, \frac{1}{2}\big), \\ \lambda^{b(1-2r)-1}, & r \in \big[\frac{1}{2}, 1\big], \\ \lambda^{-2r}, & r \in (1, \infty), \end{cases}$$

因此，(B.1) 中的近似项以至少 $1 - \delta$ 的概率满足，

$$\|G_\rho - \mathcal{S}_M F^*_\lambda\|_{L^2(\rho_\mathcal{U})} \;\le\; C_\bullet\, \lambda^r.$$

证明。 根据假设 3.2，我们有 $G_\rho = \mathcal{L}^r H$，其中 $\|H\|_{L^2(\rho_\mathcal{U})} \le R$。因此，

$$\begin{aligned}\|G_\rho - \mathcal{S}_M F_\lambda^*\|_{L^2(\rho_\mathcal{U})} &= \big\|\big(\mathcal{L}_M \phi_\lambda(\mathcal{L}_M) - I\big)\mathcal{L}^r H\big\|_{L^2(\rho_\mathcal{U})} \\ &\le R\big\|r_\lambda(\mathcal{L}_M)\,\mathcal{L}^r\big\|, \end{aligned} \tag{B.2}$$

其中 r_λ 表示式 (2.9) 中定义的残差多项式。

对于剩余项，我们有

$$\begin{aligned} R\big\|r_\lambda(\mathcal{L}_M)\,\mathcal{L}^r\big\| &\le R\big\|r_\lambda(\mathcal{L}_M)\,\mathcal{L}_{M,\lambda}^{(r\vee 1)}\big\|\big\|\mathcal{L}_{M,\lambda}^{-(r\vee 1)}\mathcal{L}_\lambda^r\big\| \\ &\le 3R\,c_{r\vee 1}\,\lambda^r, \end{aligned}$$

其中，最后一个不等式由以下界得出：$\|r_\lambda(\mathcal{L}_M)\mathcal{L}_{M,\lambda}^{(r\vee 1)}\| \le c_{r\vee 1}\lambda^{(r\vee 1)}$ 由式 (2.10) 得出，以及 $\|\mathcal{L}_{M,\lambda}^{-(r\vee 1)}\mathcal{L}_\lambda^r\| \le 3\lambda^{-(1-r)^+}$ 由命题 B.16得出；该结论以至少 $1-3\delta$ 的概率成立。 □

B.3 估计误差的界

现在我们来界定式 (B.1) 中的方差型误差。为此，我们将其分解为两部分：(i) 经典估计误差项，可通过标准集中不等式加以控制；(ii) 额外的近似型项，其贡献则利用残差多项式的性质进行界定。

命题 B.2。 假设 *3.1*、*2.1*、*3.2*和 *3.3* 均成立，并令 $\nu \ge r \vee 1$。那么，对于任意 $s \in [0, \frac{1}{2}]$ 和 $\lambda \in (0,1],$，以下结论以至少 $1-\delta$ 的概率成立：

$$\Big\|\Sigma_M^{\frac{1}{2}-s}(F_\lambda^M - F_\lambda^*)\Big\|_{\mathcal{H}_M} \le C_\bullet \log\frac{1}{\delta}\,\lambda^{r-s},$$

前提是

$$M \ge p\,C_\bullet \log^2\!\big(\delta^{-1}\big)\,\log\!\big(\lambda^{-1}\big)\cdot\begin{cases}\lambda^{-1}, & r\in\big(0,\frac{1}{2}\big),\\ \lambda^{b(1-2r)-1}, & r\in\big[\frac{1}{2},1\big],\\ \lambda^{-2r}, & r\in(1,\infty),\end{cases}$$

and

$$n \ge C_\bullet \log^{3(2r+b)}\!\big(\delta^{-1}\big)\,\lambda^{-(2r+b)}, \quad n \ge n_0 := \exp\!\Big(\tfrac{2r+b}{2r+b-1}\Big).$$

证明。 我们从分解开始

$$\begin{aligned}\big\|\Sigma_M^{\frac{1}{2}-s}(F_\lambda^M - F_\lambda^*)\big\|_{\mathcal{H}_M} &\le \big\|\Sigma_M^{\frac{1}{2}-s}\big(\phi_\lambda(\widehat{\Sigma}_M)\widehat{\mathcal{S}}_M^*\mathbf{v} - \phi_\lambda(\widehat{\Sigma}_M)\widehat{\Sigma}_M F_\lambda^*\big)\big\|_{\mathcal{H}_M} \\ &\quad + \big\|\Sigma_M^{\frac{1}{2}-s}\big(\phi_\lambda(\widehat{\Sigma}_M)\widehat{\Sigma}_M - I\big)F_\lambda^*\big\|_{\mathcal{H}_M} \\ &= \big\|\Sigma_M^{\frac{1}{2}-s}\phi_\lambda(\widehat{\Sigma}_M)\widehat{\mathcal{S}}_M^*\big(\mathbf{v} - \widehat{\mathcal{S}}_M F_\lambda^*\big)\big\|_{\mathcal{H}_M} + \big\|\Sigma_M^{\frac{1}{2}-s}r_\lambda(\widehat{\Sigma}_M)F_\lambda^*\big\|_{\mathcal{H}_M} \\ &=: \mathbf{(I)} + \mathbf{(II)}. \end{aligned} \tag{B.3}$$

我们分别对两项 **（I）**和 **（II）**进行界定。具体而言，根据命题 B.3，以至少 $1-\delta$的概率，

$$\mathbf{(I)} \le C_\bullet \log\tfrac{1}{\delta}\,\lambda^{r-s},$$

and

$$\mathbf{(II)} \le C_\bullet\,\lambda^{r-s}.$$

结合这些界即可得到所述结果。 □

命题 B.3。 假设命题 *B.2* 中的假设 $s \in [0, \frac{1}{2}]$ 和$\lambda \in (0,1]$, 的概率至少为 $1-\delta$, –

$$\textit{(I)} \quad \big\|\Sigma_M^{\frac{1}{2}-s}\,\phi_\lambda(\widehat{\Sigma}_M)\,\widehat{\mathcal{S}}_M^*\big(\mathbf{v} - \widehat{\mathcal{S}}_M F_\lambda^*\big)\big\|_{\mathcal{H}_M} \le C_\bullet \log\tfrac{1}{\delta}\,\lambda^{r-s}, \tag{B.4}$$

$$\textit{(II)} \quad \big\|\Sigma_M^{\frac{1}{2}-s}\,r_\lambda(\widehat{\Sigma}_M)\,F_\lambda^*\big\|_{\mathcal{H}_M} \le C_\bullet\,\lambda^{r-s}. \tag{B.5}$$

证明。 **(I)** 我们从分解开始

$$\begin{aligned}\left\|\Sigma_M^{\frac{1}{2}-s}\,\phi_\lambda(\widehat{\Sigma}_M)\,\widehat{\mathcal{S}}_M^*\big(\mathbf{v}-\widehat{\mathcal{S}}_M F_\lambda^*\big)\right\|_{\mathcal{H}_M} &\le \left\|\Sigma_M^{\frac{1}{2}-s}\phi_\lambda(\widehat{\Sigma}_M)\Sigma_{M,\lambda}^{\frac{1}{2}}\right\| \left\|\Sigma_{M,\lambda}^{-\frac{1}{2}}\widehat{\mathcal{S}}_M^*\big(\mathbf{v}-\widehat{\mathcal{S}}_M F_\lambda^*\big)\right\|_{\mathcal{H}_M}\\ &=: i\cdot ii. \end{aligned} \tag{B.6}$$

步骤 (i)。 根据命题 B.15，以至少 $1-4\delta$ 的概率，

$$\|\widehat{\Sigma}_{M,\lambda}^{-\frac{1}{2}}\Sigma_{M,\lambda}^{\frac{1}{2}}\| \le 2. \tag{B.7}$$

因此

$$\begin{aligned}\left\|\Sigma_M^{\frac{1}{2}-s}\phi_\lambda(\widehat{\Sigma}_M)\Sigma_{M,\lambda}^{\frac{1}{2}}\right\| &\le \lambda^{-s}\left\|\Sigma_{M,\lambda}^{\frac{1}{2}}\phi_\lambda(\widehat{\Sigma}_M)\Sigma_{M,\lambda}^{\frac{1}{2}}\right\|\\ &\le \lambda^{-s}\left\|\widehat{\Sigma}_{M,\lambda}\phi_\lambda(\widehat{\Sigma}_M)\right\| \left\|\Sigma_{M,\lambda}^{\frac{1}{2}}\widehat{\Sigma}_{M,\lambda}^{-\frac{1}{2}}\right\|^2 \;\le\; 4D\,\lambda^{-s},\end{aligned}$$

其中 D 定义见（2.7）。

步骤 (ii)。我们将其分解为

$$\begin{aligned}\left\|\Sigma_{M,\lambda}^{-\frac{1}{2}}\widehat{\mathcal{S}}_M^*\big(\mathbf{v}-\widehat{\mathcal{S}}_M F_\lambda^*\big)\right\|_{\mathcal{H}_M} &\le \left\|\Sigma_{M,\lambda}^{-\frac{1}{2}}\widehat{\mathcal{S}}_M^*\big(\mathbf{v}-\bar{G}_\rho\big)\right\|_{\mathcal{H}_M} + \left\|\Sigma_{M,\lambda}^{-\frac{1}{2}}\widehat{\mathcal{S}}_M^*\big(\bar{G}_\rho-\widehat{\mathcal{S}}_M F_\lambda^*\big)\right\|_{\mathcal{H}_M}\\ &=: a+b.\end{aligned}$$

项 ***(a)***：结合命题 B.22 与命题 B.18，我们得到，以至少$1-3\delta$ 的概率，

$$\begin{aligned}\left\|\Sigma_{M,\lambda}^{-\frac{1}{2}}\widehat{\mathcal{S}}_M^*\big(\mathbf{v}-\bar{G}_\rho\big)\right\|_{\mathcal{H}_M} &\le \left(\frac{4QZ\kappa}{\sqrt{\lambda}n}+\frac{8Q\sqrt{(1+2\log\frac{2}{\delta})\mathcal{N}_{\mathcal{L}}(\lambda)}}{\sqrt{n}}\right)\log\tfrac{2}{\delta}\\ &\le C_\bullet\left(\frac{1}{\sqrt{\lambda}n}+\sqrt{\frac{\log\frac{1}{\delta}}{n\lambda^b}}\right)\log\tfrac{1}{\delta} \;\le\; C_\bullet\,\lambda^r\log\tfrac{1}{\delta},\end{aligned}$$

其中，最后一个不等式由关于 n 的假设推出。项 ***(b)***：根据命题B.15，

$$\left\|\Sigma_{M,\lambda}^{-\frac{1}{2}}\widehat{\mathcal{S}}_M^*\right\|^2 \le 2\|\widehat{\Sigma}_{M,\lambda}^{-\frac{1}{2}}\widehat{\mathcal{S}}_M^*\|^2 = 2\left\|\widehat{\mathcal{S}}_M^*\widehat{\mathcal{S}}_M(\widehat{\mathcal{S}}_M^*\widehat{\mathcal{S}}_M+\lambda)^{-1}\right\| \le 2.$$

因此，根据命题 B.23，我们以至少 $1-\delta$ 的概率得到，

$$\begin{aligned}\left\|\Sigma_{M,\lambda}^{-\frac{1}{2}}\widehat{\mathcal{S}}_M^*(\bar{G}_\rho-\widehat{\mathcal{S}}_M F_\lambda^*)\right\|_{\mathcal{H}_M} &\le \frac{2}{\sqrt{n}}\|\bar{G}_\rho-\widehat{\mathcal{S}}_M F_\lambda^*\|_2\\ &\le 2\sqrt{\left|\frac{1}{n}\|\bar{G}_\rho-\widehat{\mathcal{S}}_M F_\lambda^*\|_2^2-\|G_\rho-\mathcal{S}_M F_\lambda^*\|_{L^2(\rho_{\mathcal{U}})}^2\right|}+\|G_\rho-\mathcal{S}_M F_\lambda^*\|_{L^2(\rho_{\mathcal{U}})}\\ &\le C_\bullet\sqrt{\left(\frac{\lambda^{-2(\frac{1}{2}-r)^+}}{n}+\frac{\lambda^{-(\frac{1}{2}-r)^+}\|G_\rho-\mathcal{S}_M F_\lambda^*\|_{L^2(\rho_{\mathcal{U}})}}{\sqrt{n}}\right)\log\tfrac{1}{\delta}}+\|G_\rho-\mathcal{S}_M F_\lambda^*\|_{L^2(\rho_{\mathcal{U}})}.\end{aligned}$$

应用命题 B.1 可得，以至少 $1-3\delta$ 的概率，

$$\begin{aligned}\left\|\Sigma_{M,\lambda}^{-\frac{1}{2}}\widehat{\mathcal{S}}_M^*(\bar{G}_\rho-\widehat{\mathcal{S}}_M F_\lambda^*)\right\|_{\mathcal{H}_M} &\le C_\bullet\sqrt{\left(\frac{\lambda^{-(1-2r)^+}}{n}+\frac{\lambda^{-(\frac{1}{2}-r)^+}\lambda^r}{\sqrt{n}}+\lambda^{2r}\right)\log\tfrac{1}{\delta}}\\ &\le C_\bullet\,\lambda^r\sqrt{\log\tfrac{1}{\delta}},\end{aligned}$$

其中，最后一个不等式由关于 n的假设推出。因此，

$$ii \le C_\bullet\,\lambda^r\log\tfrac{1}{\delta}.$$

将（i）和（ii）代入（B.6）即可证明（B.4）。汇总所有概率并应用命题 B.5，可得总概率至少为 $1-11\delta$。重新定义 δ 即可完成第（I）部分。

（II）接下来，我们对式（B.5）中的项进行界定。以至少 $1-4\delta$ 的概率，

$$\begin{aligned}\|\Sigma_M^{\frac{1}{2}-s} r_\lambda(\widehat{\Sigma}_M) F_\lambda^*\|_{\mathcal{H}_M} &\le \lambda^{-s}\|\Sigma_{M,\lambda}^{\frac{1}{2}} r_\lambda(\widehat{\Sigma}_M) F_\lambda^*\|_{\mathcal{H}_M} \\ &\le \lambda^{-s}\|\Sigma_{M,\lambda}^{\frac{1}{2}}\widehat{\Sigma}_{M,\lambda}^{-\frac{1}{2}}\|\,\|\widehat{\Sigma}_{M,\lambda}^{\frac{1}{2}} r_\lambda(\widehat{\Sigma}_M) F_\lambda^*\|_{\mathcal{H}_M} \\ &\le 2\lambda^{-s}\|\widehat{\Sigma}_{M,\lambda}^{\frac{1}{2}} r_\lambda(\widehat{\Sigma}_M) F_\lambda^*\|_{\mathcal{H}_M},\end{aligned}$$

这里我们再次使用了命题 B.15。展开 $F_\lambda^* = \mathcal{S}_M^* \phi_\lambda(\mathcal{L}_M)\mathcal{L}^r H$ 可得

$$2\lambda^{-s}\|\widehat{\Sigma}_{M,\lambda}^{\frac{1}{2}} r_\lambda(\widehat{\Sigma}_M) F_\lambda^*\|_{\mathcal{H}_M} \;\le\; 2R\lambda^{-s}\|\widehat{\Sigma}_{M,\lambda}^{\frac{1}{2}} r_\lambda(\widehat{\Sigma}_M)\mathcal{S}_M^*\phi_\lambda(\mathcal{L}_M)\mathcal{L}^r\|. \tag{B.8}$$

我们区分两种情形。

情形 $r \le \frac{1}{2}$。我们有

$$\begin{aligned}\|\widehat{\Sigma}_{M,\lambda}^{\frac{1}{2}} r_\lambda(\widehat{\Sigma}_M)\mathcal{S}_M^*\phi_\lambda(\mathcal{L}_M)\mathcal{L}^r\| &\le \|\widehat{\Sigma}_{M,\lambda}^{\frac{1}{2}} r_\lambda(\widehat{\Sigma}_M)\mathcal{S}_M^*\phi_\lambda(\mathcal{L}_M)\mathcal{L}_{M,\lambda}^r\|\|\mathcal{L}_{M,\lambda}^{-r}\mathcal{L}_\lambda^r\| \\ &= \|\widehat{\Sigma}_{M,\lambda}^{\frac{1}{2}} r_\lambda(\widehat{\Sigma}_M)\Sigma_{M,\lambda}^r\mathcal{S}_M^*\phi_\lambda(\mathcal{L}_M)\|\|\mathcal{L}_{M,\lambda}^{-r}\mathcal{L}_\lambda^r\| \\ &\le \|\widehat{\Sigma}_{M,\lambda}^{\frac{1}{2}} r_\lambda(\widehat{\Sigma}_M)\Sigma_{M,\lambda}^r\|\|\mathcal{L}_M^{\frac{1}{2}}\phi_\lambda(\mathcal{L}_M)\|\|\mathcal{L}_{M,\lambda}^{-r}\mathcal{L}_\lambda^r\|.\end{aligned}$$

根据命题 B.10，$\|\mathcal{L}_M^{\frac{1}{2}}\phi_\lambda(\mathcal{L}_M)\| \le D\lambda^{-1/2}$。此外，根据命题 B.7 和命题 B.14，$\|\mathcal{L}_{M,\lambda}^{-r}\mathcal{L}_\lambda^r\| \le 2$ 以至少 $1-4\delta$ 的概率成立。因此，

$$\|\Sigma_M^{\frac{1}{2}-s} r_\lambda(\widehat{\Sigma}_M) F_\lambda^*\|_{\mathcal{H}_M} \le 4DR\lambda^{-s-\frac{1}{2}}\|\widehat{\Sigma}_{M,\lambda}^{\frac{1}{2}} r_\lambda(\widehat{\Sigma}_M)\Sigma_{M,\lambda}^r\|. \tag{B.9}$$

$\widehat{}$为界定剩余项，使用命题 B.15：$\|\Sigma_{M,\lambda}^{-r}\Sigma_{M,\lambda}^r\| \le 2$。结合式（2.10），可得

$$\|\widehat{\Sigma}_{M,\lambda}^{\frac{1}{2}} r_\lambda(\widehat{\Sigma}_M)\Sigma_{M,\lambda}^r\| \le 2c_{\frac{1}{2}+r}\lambda^{\frac{1}{2}+r}.$$

将其代入（B.9）得

$$\|\Sigma_M^{\frac{1}{2}-s} r_\lambda(\widehat{\Sigma}_M) F_\lambda^*\|_{\mathcal{H}_M} \le 8DR\, c_{\frac{1}{2}+r}\lambda^{r-s}.$$

情形 $r > \frac{1}{2}$。我们采用类似的方法：

$$\begin{aligned}\|\widehat{\Sigma}_{M,\lambda}^{\frac{1}{2}} r_\lambda(\widehat{\Sigma}_M)\mathcal{S}_M^*\phi_\lambda(\mathcal{L}_M)\mathcal{L}^r\| &\le \|\widehat{\Sigma}_{M,\lambda}^{\frac{1}{2}} r_\lambda(\widehat{\Sigma}_M)\mathcal{S}_M^*\phi_\lambda(\mathcal{L}_M)\mathcal{L}_{M,\lambda}^{(r\vee 1)}\|\|\mathcal{L}_{M,\lambda}^{-(r\vee 1)}\mathcal{L}_\lambda^r\| \\ &\le \|\widehat{\Sigma}_{M,\lambda}^{\frac{1}{2}} r_\lambda(\widehat{\Sigma}_M)\Sigma_{M,\lambda}^{(r\vee 1)-\frac{1}{2}}\|\|\mathcal{L}_M\phi_\lambda(\mathcal{L}_M)\|\|\mathcal{L}_{M,\lambda}^{-(r\vee 1)}\mathcal{L}_\lambda^r\|.\end{aligned}$$

根据谱方法的性质和命题 B.16，$\|\mathcal{L}_M\phi_\lambda(\mathcal{L}_M)\| \le D$ 和 $\|\mathcal{L}_{M,\lambda}^{-(r\vee 1)}\mathcal{L}_\lambda^r\| \le 3\lambda^{-(1-r)^+}$。因此，

$$\|\Sigma_M^{\frac{1}{2}-s} r_\lambda(\widehat{\Sigma}_M) F_\lambda^*\|_{\mathcal{H}_M} \le \frac{6DR}{\lambda^{s+(1-r)^+}}\|\widehat{\Sigma}_{M,\lambda}^{\frac{1}{2}} r_\lambda(\widehat{\Sigma}_M)\Sigma_{M,\lambda}^{(r\vee 1)-\frac{1}{2}}\|. \tag{B.10}$$

利用（2.10）和命题 B.17，概率至少为 $1-\delta$，

$$\|\widehat{\Sigma}_{M,\lambda}^{\frac{1}{2}} r_\lambda(\widehat{\Sigma}_M)\Sigma_{M,\lambda}^{(r\vee 1)-\frac{1}{2}}\| \le 2c_{r\vee 1}\lambda^{(r\vee 1)}. \tag{B.11}$$

将其代入（B.10）得

$$\|\Sigma_M^{\frac{1}{2}-s} r_\lambda(\widehat{\Sigma}_M) F_\lambda^*\|_{\mathcal{H}_M} \le 12DR\, c_{\frac{1}{2}+r}\lambda^{r-s}.$$

结合这两种情形即可得到（B.5）。汇总所有集中界并应用命题 B.5，可得概率至少为 $1-8\delta$。重新定义 δ 即可完成证明。 □

B.4 合并误差界

定理 B.4。 假设 *3.1*、*2.1*、*3.2*和 *3.3* 成立，且令 $\nu \geq r \vee 1$。那么，对于任意$s \in [0, \frac{1}{2}]$和 $\lambda \in (0, 1],$，以下结论以至少 $1-\delta$的概率成立：

$$\|G_\rho - \mathcal{S}_M F_\lambda^M\|_{L^2(\rho_\mathcal{U})} \leq C_\bullet \log\tfrac{1}{\delta}\, \lambda^r,$$

前提是

$$M \geq p C_\bullet \log^2\!\big(\delta^{-1}\big) \log\!\big(\lambda^{-1}\big) \cdot \begin{cases} \lambda^{-1}, & r \in \big(0, \frac{1}{2}\big), \\ \lambda^{b(1-2r)-1}, & r \in \big[\frac{1}{2}, 1\big], \\ \lambda^{-2r}, & r \in (1, \infty), \end{cases}$$

and

$$n \geq C_\bullet \log^{3(2r+b)}\!\big(\delta^{-1}\big)\, \lambda^{-(2r+b)}, \quad n \geq n_0 := \exp\!\Big(\tfrac{2r+b}{2r+b-1}\Big).$$

证明。 我们从以下分解开始：

$$\|G_\rho - \mathcal{S}_M F_\lambda^M\|_{L^2(\rho_\mathcal{U})} \leq \|G_\rho - \mathcal{S}_M F_\lambda^*\|_{L^2(\rho_\mathcal{U})} + \|\mathcal{S}_M(F_\lambda^M - F_\lambda^*)\|_{L^2(\rho_\mathcal{U})} =: T_1 + T_2. \tag{B.12}$$

我们现在分别界定 T_1 和 T_2 。

步骤 1：界定 T_1。 根据命题 B.1，以至少 $1-\delta$ 的概率，

$$\|G_\rho - \mathcal{S}_M F_\lambda^*\|_{L^2(\rho_\mathcal{U})} \leq C_\bullet \lambda^r. \tag{B.13}$$

步骤 2：界定 T_2。 根据 Mercer 定理（例如，参见 Steinwart 和 Christmann (2008年）和命题 B.2，以至少 $1-\delta$ 的概率，

$$\|\mathcal{S}_M(F_\lambda^M - F_\lambda^*)\|_{L^2(\rho_\mathcal{U})} = \|\Sigma_M^{\frac{1}{2}}(F_\lambda^M - F_\lambda^*)\|_{\mathcal{H}_M} \leq C_\bullet \log\tfrac{1}{\delta}\, \lambda^r.$$

结论。 结合 T_1 和 T_2 在（B.12）中的界，即可得到所需结果：

$$\|G_\rho - \mathcal{S}_M F_\lambda^M\|_{L^2(\rho_\mathcal{U})} \leq C_\bullet \log\tfrac{1}{\delta}\, \lambda^r.$$

□

最后指出，定理3.4 可直接由定理B.4推出。

B.5 技术不等式

命题 B.5。 设 E_i 为概率至少为 $1-\delta_i$ 的事件，并令

$$E := \bigcap_{i=1}^k E_i.$$

若能证明对某个事件 A 有 $\mathbb{P}(A|E) \geq 1-\delta$，则也有

$$\begin{aligned} \mathbb{P}(A) &\geq \int_E \mathbb{P}(A|\omega) d\mathbb{P}(\omega) \geq (1-\delta)\mathbb{P}(E) \\ &= (1-\delta)\left(1 - \mathbb{P}\left(\bigcup_{i=1}^k (\Omega/E_i)\right)\right) \geq (1-\delta)\left(1 - \sum_{i=1}^k \delta_i\right). \end{aligned}$$

命题 B.6 (Aleksandrov 和 Peller(2009)， Blanchard 和 M¨ucke (2017)（命题 B.1。）)。设 B_1， B_2 为某个希尔伯特空间上的两个非负自伴算子，且 $\|B_j\| \leq a$， $j = 1, 2$，其中 a 为某个非负数。

(i) If $0 \leq r \leq 1$, *then*

$$\|B_1^r - B_2^r\| \leq C_r \|B_1 - B_2\|^r,$$

对于某个 $C_r < \infty$。

(ii) 若 $r > 1$，则

$$\|B_1^r - B_2^r\| \le C_{a,r}\|B_1 - B_2\|,$$

对某些 $C_{a,r} < \infty$。

命题 B.7（Fujii 等人，1993年，Cordes 不等式）。设 A 和 B 是可分希尔伯特空间上的两个正有界线性算子。那么

$$\|A^s B^s\| \le \|AB\|^s, \quad when\ 0 \le s \le 1.$$

命题 B.8（Rudi 和 Rosasco（2016年）（命题 9）。设 $\mathcal{H}$，$\mathcal{K}$ 是两个可分希尔伯特空间，且 A，B 有界线性算子，其中 $A:\ \mathcal{H} \to \mathcal{K}$ 和 $B:\ \mathcal{H} \to \mathcal{H}$ 是半正定的。

$$\|AB^\sigma\| \le \|A\|^{1-\sigma}\|AB\|^\sigma, \quad \forall \sigma \in [0,1].$$

命题 B.9. 设 H_1、H_2 为两个可分希尔伯特空间，且 $\mathcal{S}:\ H_1 \to H_2$ 为一个紧算子。则对任意函数 $f:$ $[0, \|\mathcal{S}\|] \to [0, \infty[$,

$$f(\mathcal{S}\mathcal{S}^*)\mathcal{S} = \mathcal{S} f(\mathcal{S}^*\mathcal{S}).$$

证明。 该结果可利用紧算子的奇异值分解加以证明。 □

命题 B.10（Lin 和 Cevher（2018）（引理 10））。设 L 为可分希尔伯特空间 H 上的紧的正算子，使得 $\|L\| \le \kappa^2$。则对任意 $\lambda \ge 0$,

$$\|(L+\lambda)^\alpha \phi_\lambda(L)\| \le 2D\lambda^{-(1-\alpha)}, \quad \forall \alpha \in [0,1],$$
$$\|L^\alpha \phi_\lambda(L)\| \le D\lambda^{-(1-\alpha)}, \quad \forall \alpha \in [0,1],$$

其中 D 定义于（2.7）。

命题 B.11. 以至少为 $1-\delta$ 的概率，下列界成立:

$$\|F_\lambda^*\|_\infty \le 2\,\kappa^{2r+1}\,R\,D\,\lambda^{-(\frac{1}{2}-r)^+},$$
$$\|F_\lambda^*\|_{\mathcal{H}_M} \le 2\,\kappa^{2r}\,R\,D\,\lambda^{-(\frac{1}{2}-r)^+},$$

前提是

$$M \ \ge\ \frac{8\,p\,\kappa^2\,\beta_\infty}{\lambda}, \quad with \quad \beta_\infty = \log\frac{4\,\kappa^2\big(\mathcal{N}_{\mathcal{L}}(\lambda)+1\big)}{\delta\,\|\mathcal{L}\|}.$$

证明。 由于 $F_\lambda^* \in \mathcal{H}_M$，根据再生性质和定义 $G_\rho = \mathcal{L}^r H$，我们得到，对于任意 $x \in \mathcal{X}$,

$$\|F_\lambda^*(x)\|_{\mathcal{Y}} = \|K_{M,x}^* F_\lambda^*\|_{\mathcal{Y}} \ \le\ \kappa\,\|F_\lambda^*\|_{\mathcal{H}_M} = \kappa\,\|\mathcal{S}_M^*\phi_\lambda(\mathcal{L}_M)G_\rho\|_{\mathcal{H}_M}$$
$$= \kappa\,\|\mathcal{L}_M^{\frac{1}{2}}\phi_\lambda(\mathcal{L}_M)\mathcal{L}^r H\|_{L^2(\rho_x)}.$$

利用 $\|H\|_{L^2(\rho_x)} \le R$，我们得到

$$\|F_\lambda^*\|_\infty \le \kappa\,R\,\|\mathcal{L}_M^{\frac{1}{2}}\phi_\lambda(\mathcal{L}_M)\mathcal{L}_{M,\lambda}^{(r\wedge\frac{1}{2})}\|\,\|\mathcal{L}_{M,\lambda}^{-(r\wedge\frac{1}{2})}\mathcal{L}^r\| = \kappa\,R\,(I)\,(II), \tag{B.14}$$

$$\|F_\lambda^*\|_{\mathcal{H}_M} \le R\,\|\mathcal{L}_M^{\frac{1}{2}}\phi_\lambda(\mathcal{L}_M)\mathcal{L}_{M,\lambda}^{(r\wedge\frac{1}{2})}\|\,\|\mathcal{L}_{M,\lambda}^{-(r\wedge\frac{1}{2})}\mathcal{L}^r\| = R\,(I)\,(II). \tag{B.15}$$

步骤（I）。根据命题B.10,

$$I = \big\|\mathcal{L}_M^{\frac{1}{2}+(r\wedge\frac{1}{2})}\phi_\lambda(\mathcal{L}_M)\big\| \le \begin{cases} D, & r \ge \frac{1}{2}, \\ D\,\lambda^{r-\frac{1}{2}}, & r < \frac{1}{2}, \end{cases}$$
$$\le D\,\lambda^{-(\frac{1}{2}-r)^+}.$$

步骤（II）。根据命题 B.7 和 B.14，以至少为 $1-\delta$的概率，我们得到

$$II = \begin{cases} \|\mathcal{L}_{M,\lambda}^{-\frac{1}{2}}\mathcal{L}_\lambda^r\| \le \|\mathcal{L}_{M,\lambda}^{-\frac{1}{2}}\mathcal{L}_\lambda^{\frac{1}{2}}\|\|\mathcal{L}^{r-\frac{1}{2}}\| \le 2\,\kappa^{2r-1}, & r \ge \frac{1}{2}, \\ \|\mathcal{L}_{M,\lambda}^{-r}\mathcal{L}_\lambda^r\| \le \|\mathcal{L}_{M,\lambda}^{-\frac{1}{2}}\mathcal{L}_\lambda^{\frac{1}{2}}\|^{2r} \le 4^r \le 2, & r < \frac{1}{2}, \end{cases}$$
$$\le 2\,\kappa^{2r}.$$

结合 I 和 II 在（B.14）和（B.15中的界可得

$$\|F_\lambda^*\|_\infty \le 2\,\kappa^{2r+1}\,R\,D\,\lambda^{-(\frac{1}{2}-r)^+},$$
$$\|F_\lambda^*\|_{\mathcal{H}_M} \le 2\,\kappa^{2r}\,R\,D\,\lambda^{-(\frac{1}{2}-r)^+}.$$

□

命题 B.12。 设 $\mathcal{H}$ 是一个可分希尔伯特空间，且 A 和 B 是作用于 $\mathcal{H}$ 和 $\lambda > 0$ 上的两个有界自伴正线性算子。则

$$\left\|A_\lambda^{-\frac{1}{2}}B_\lambda^{\frac{1}{2}}\right\| \le (1-c)^{-\frac{1}{2}}, \quad \left\|A_\lambda^{\frac{1}{2}}B_\lambda^{-\frac{1}{2}}\right\| \le (1+c)^{\frac{1}{2}},$$

with

$$c = \left\|B_\lambda^{-\frac{1}{2}}(A-B)B_\lambda^{-\frac{1}{2}}\right\|.$$

证明。 第一个不等式的证明例如可见 Rudi 和 Rosasco （2016年）（命题 8）。通过简单计算可得第二个不等式:

$$\begin{aligned}\left\|(A+\lambda I)^{\frac{1}{2}}(B+\lambda I)^{-\frac{1}{2}}\right\|^2 &= \left\|(B+\lambda I)^{-\frac{1}{2}}(A+\lambda I)(B+\lambda I)^{-\frac{1}{2}}\right\| \\ &\le \left\|(B+\lambda I)^{-\frac{1}{2}}(A-B)(B+\lambda I)^{-\frac{1}{2}}\right\| + \|I\| \le 1+c.\end{aligned}$$

□

命题 B.13 （Rudi 和 Rosasco （2016年）（引理 9））。 对于任意 $M \ge 8\kappa^4\|\mathcal{L}\|^{-1}\log^2\frac{2}{\delta}$，我们以至少 $1-\delta$ 的概率有

$$\|\mathcal{L}_M\| \ge \frac{1}{2}\|\mathcal{L}\|.$$

证明。 对于 $M \ge 8\kappa^4\|\mathcal{L}\|^{-1}\log^2\frac{2}{\delta}$，根据命题 B.21 （$E_6$），以至少 $1-\delta$ 的概率，
$\|\mathcal{L}-\mathcal{L}_M\|_{HS} \le \frac{1}{2}\|\mathcal{L}\|$ 因此

$$\|\mathcal{L}_M\| \ge \|\mathcal{L}\| - \|\mathcal{L}-\mathcal{L}_M\|_{HS} \ge \frac{1}{2}\|\mathcal{L}\|.$$

□

命题 B.14。 根据假设 *2.1* ，对于任意 $M \ge \frac{8p\kappa^2\beta_\infty}{\lambda}$，其中 $\beta_\infty = \log\frac{4\kappa^2(\mathcal{N}_\mathcal{L}(\lambda)+1)}{\delta\|\mathcal{L}\|}$，~~其概率至少为~~ $1-\delta$

$$\left\|\mathcal{L}_{M,\lambda}^{-\frac{1}{2}}\mathcal{L}_\lambda^{\frac{1}{2}}\right\| \le 2, \quad \left\|\mathcal{L}_{M,\lambda}^{\frac{1}{2}}\mathcal{L}_\lambda^{-\frac{1}{2}}\right\| \le 2.$$

证明。 由命题 B.21 (E_2)，对于任意 $\lambda > 0$,

$$\left\|\mathcal{L}_\lambda^{-\frac{1}{2}}(\mathcal{L}_M-\mathcal{L})\mathcal{L}_\lambda^{-\frac{1}{2}}\right\| \le \frac{4\kappa^2\beta_\infty}{3M\lambda} + \sqrt{\frac{2p\kappa^2\beta_\infty}{M\lambda}}. \tag{B.16}$$

由 $M \ge \frac{8p\kappa^2\beta_\infty}{\lambda}$，因此得到

$$\left\|\mathcal{L}_\lambda^{-\frac{1}{2}}(\mathcal{L}_M-\mathcal{L})\mathcal{L}_\lambda^{-\frac{1}{2}}\right\| \le \frac{3}{4}. \tag{B.17}$$

结论现在可由命题 B.12 □

命题 B.15. 在满足假设 *2.1* 的情况下，对任意 $n \geq \frac{8\kappa^2\tilde{\beta}}{\lambda}$ 满足 $\tilde{\beta} \geq \log\frac{4\kappa^2(\mathcal{N}_{\mathcal{L}}(\lambda)+1)}{\delta\|\mathcal{L}\|}$ 且 $M \geq \frac{8p\kappa^2\beta_\infty}{\lambda} \vee 8\kappa^4\|\mathcal{L}\|^{-1}\log^2\frac{2}{\delta}$，其中 $\beta_\infty = \log\frac{4\kappa^2(\mathcal{N}_{\mathcal{L}}(\lambda)+1)}{\delta\|\mathcal{L}\|}$ 从而以至少 $1-4\delta$ 的概率

$$\left\|\widehat{\Sigma}_{M,\lambda}^{-\frac{1}{2}}\Sigma_{M,\lambda}^{\frac{1}{2}}\right\| \leq 2, \quad \left\|\widehat{\Sigma}_{M,\lambda}^{\frac{1}{2}}\Sigma_{M,\lambda}^{-\frac{1}{2}}\right\| \leq 2.$$

注意，我们将此界用于（B.7）中的方差项。此外，经简短计算可知，上述关于 n的假设在命题 B.2中所述的样本量条件下确实成立，即

$$n \geq C_\bullet \log^{3(2r+b)}\left(\delta^{-1}\right)\lambda^{-(2r+b)}, \qquad n \geq n_0 := \exp\left(\tfrac{2r+b}{2r+b-1}\right).$$

证明。由命题 B.21 (E_1)，对任意 $\lambda > 0$，以至少 $1-\delta$的概率，

$$\left\|\Sigma_{M,\lambda}^{-\frac{1}{2}}\left(\widehat{\Sigma}_M - \Sigma_M\right)\Sigma_{M,\lambda}^{-\frac{1}{2}}\right\| \leq \frac{4\kappa^2\beta_M}{3n\lambda} + \sqrt{\frac{2\kappa^2\beta_M}{n\lambda}}, \tag{B.18}$$

其中 $\beta_M = \log\frac{4\kappa^2(\mathcal{N}_{\mathcal{L}_M}(\lambda)+1)}{\delta\|\mathcal{L}_M\|}$。对于 $M \geq \frac{8p\kappa^2\beta_\infty}{\lambda}$，由命题 B.18 可得，以至少 $1-2\delta$的概率，

$$\mathcal{N}_{\mathcal{L}_M}(\lambda) \leq \left(1 + 2\log\frac{2}{\delta}\right)4\mathcal{N}_{\mathcal{L}}(\lambda). \tag{B.19}$$

由命题 B.13 我们以概率 $1-\delta$ 有，

$$\|\mathcal{L}_M\| \geq \frac{1}{2}\|\mathcal{L}\|. \tag{B.20}$$

注意，(B.19) 与 (B.20) 的界蕴含 $\beta_M \leq \tilde{\beta} = \log\frac{4\kappa^2((1+2\log\frac{2}{\delta})4\mathcal{N}_{\mathcal{L}}(\lambda)+1)}{\delta\|\mathcal{L}\|}$。将其与 $n \geq \frac{8\kappa^2\tilde{\beta}}{\lambda}$ 结合使用，我们对于 (B.18) 得到

$$\left\|\Sigma_{M,\lambda}^{-\frac{1}{2}}\left(\widehat{\Sigma}_M - \Sigma_M\right)\Sigma_{M,\lambda}^{-\frac{1}{2}}\right\| \leq \frac{4\kappa^2\beta_M}{3n\lambda} + \sqrt{\frac{2\kappa^2\beta_M}{n\lambda}} \tag{B.21}$$

$$\leq \frac{4\kappa^2\tilde{\beta}}{3n\lambda} + \sqrt{\frac{2\kappa^2\tilde{\beta}}{n\lambda}} \leq \frac{3}{4}. \tag{B.22}$$

注意到 from 命题 B.5 我们有上述不等式成立，其概率至少为 $1-4\delta$.
该结果现在由命题B.12推出 □

命题 B.16. 根据假设 *2.1*，对于任意

$$M \geq \begin{cases} \frac{8p\kappa^2\beta_\infty}{\lambda} & r \in \left(0, \frac{1}{2}\right) \\ \frac{(8p\kappa^2\beta_\infty)\vee C_1^{\frac{1}{r}}}{\lambda} \vee \frac{C_2}{\lambda^{1+b(2r-1)}} & r \in \left[\frac{1}{2}, 1\right] \\ \frac{C_3}{\lambda^{2r}} & r \in (1, \infty) \end{cases}$$

以至少 $1-3\delta$ 的概率，

$$\left\|\mathcal{L}_{M,\lambda}^{-(r\vee 1)}\mathcal{L}_\lambda^r\right\| \leq \frac{3}{\lambda^{(1-r)^+}},$$

其中 $C_1 = 2(4\kappa\log\frac{2}{\delta})^{2r-1}(8p\kappa^2\beta_\infty)^{1-r}$，$C_2 = 4(4c_b\kappa^2\log\frac{2}{\delta})^{2r-1}(8p\kappa^2\beta_\infty)^{2-2r}$，$C_3 := 4\kappa^4 C_{\kappa,r}^2\log^2\frac{2}{\delta}$ 且有 $C_{\kappa,r}$ 根据命题 *B.6*。

证明。 证明中，我们需要区分以下三种情形：

- 情形 ($r \leq \frac{1}{2}$)：根据命题 B.14 我们以至少 $1-\delta$ 的概率，

$$\begin{aligned}\left\|\mathcal{L}_{M,\lambda}^{-(r\vee 1)}\mathcal{L}_\lambda^r\right\| &= \left\|\mathcal{L}_{M,\lambda}^{-1}\mathcal{L}_\lambda^r\right\| \\ &\leq \lambda^{r-1}\left\|\mathcal{L}_{M,\lambda}^{-r}\mathcal{L}_\lambda^r\right\| \\ &\leq \lambda^{r-1}\left\|\mathcal{L}_{M,\lambda}^{-\frac{1}{2}}\mathcal{L}_\lambda^{\frac{1}{2}}\right\|^{2r} \leq 2^{2r}\lambda^{r-1} \leq 3\lambda^{r-1}.\end{aligned}$$

• 情形 ($r \in [\frac{1}{2}, 1]$)：根据 $\left\|\mathcal{L}_\lambda^{-1}\mathcal{L}_\lambda^r\right\| \le \lambda^{r-1}$ 我们有

$$\left\|\mathcal{L}_{M,\lambda}^{-(r\vee 1)}\mathcal{L}_\lambda^r\right\| = \left\|\mathcal{L}_{M,\lambda}^{-1}\mathcal{L}_\lambda^r\right\| \tag{B.23}$$

$$\le \left\|\left(\mathcal{L}_{M,\lambda}^{-1} - \mathcal{L}_\lambda^{-1}\right)\mathcal{L}_\lambda^r\right\| + \lambda^{r-1}. \tag{B.24}$$

对于最后一个不等式中的范数，根据代数恒等式 $A^{-1} - B^{-1} = A^{-1}(A-B)B^{-1}$，我们有：

$$\left\|\left(\mathcal{L}_{M,\lambda}^{-1} - \mathcal{L}_\lambda^{-1}\right)\mathcal{L}_\lambda^r\right\| = \left\|\mathcal{L}_{M,\lambda}^{-1}(\mathcal{L}_{M,\lambda} - \mathcal{L}_\lambda)\mathcal{L}_\lambda^{r-1}\right\|$$

并且根据命题 B.14 以至少 $1-\delta$ 的概率，

$$\begin{aligned}&\left\|\mathcal{L}_{M,\lambda}^{-1}(\mathcal{L}_{M,\lambda} - \mathcal{L}_\lambda)\mathcal{L}_\lambda^{r-1}\right\| \\ &\le \lambda^{-\frac{1}{2}}\left\|\mathcal{L}_{M,\lambda}^{-\frac{1}{2}}\mathcal{L}_\lambda^{\frac{1}{2}}\right\|\left\|\mathcal{L}_\lambda^{-\frac{1}{2}}(\mathcal{L}_{M,\lambda} - \mathcal{L}_\lambda)\mathcal{L}_\lambda^{r-1}\right\| \\ &\le 2\lambda^{-\frac{1}{2}}\left\|\mathcal{L}_\lambda^{-\frac{1}{2}}(\mathcal{L}_{M,\lambda} - \mathcal{L}_\lambda)\mathcal{L}_\lambda^{r-1}\right\|.\end{aligned}$$

由于 $\sigma := 2 - 2r \le 1$ 根据命题 B.8

$$\begin{aligned}&\left\|\mathcal{L}_\lambda^{-\frac{1}{2}}(\mathcal{L}_{M,\lambda} - \mathcal{L}_\lambda)\mathcal{L}_\lambda^{r-1}\right\| \\ &\le \left\|\mathcal{L}_\lambda^{-\frac{1}{2}}(\mathcal{L}_{M,\lambda} - \mathcal{L}_\lambda)\right\|^{2r-1}\left\|\mathcal{L}_\lambda^{-\frac{1}{2}}(\mathcal{L}_{M,\lambda} - \mathcal{L}_\lambda)\mathcal{L}_\lambda^{-\frac{1}{2}}\right\|^{2-2r}.\end{aligned}$$

利用 命题 B.21 (E_2 和 E_5)，对于最后一个表达式，以至少 $1-2$ 的概率有 δ,

$$\le \left[\left(\frac{2\kappa}{\sqrt{\lambda}M} + \sqrt{\frac{4\kappa^2\mathcal{N}_\mathcal{L}(\lambda)}{M}}\right)\log\frac{2}{\delta}\right]^{2r-1}\left(\frac{4\kappa^2\beta_\infty}{3M\lambda} + \sqrt{\frac{2p\kappa^2\beta_\infty}{M\lambda}}\right)^{2-2r},$$

其中 $\beta_\infty = \log 4\kappa^2(\mathcal{N}_\mathcal{L}(\lambda)+1)/\delta\|\mathcal{L}\|$。结合这一结果与 $M \ge \frac{8p\kappa^2\beta_\infty}{\lambda}$ 以及简单不等式 $(a+b)^{2r-1} \le a^{2r-1} + b^{2r-1}$，我们得到

$$\begin{aligned}&\left\|\left(\mathcal{L}_{M,\lambda}^{-1} - \mathcal{L}_\lambda^{-1}\right)\mathcal{L}_\lambda^r\right\| \\ &\le 2\lambda^{-\frac{1}{2}}\left(\frac{4\kappa\log\frac{2}{\delta}}{\sqrt{\lambda}M} + \sqrt{\frac{4\kappa^2\mathcal{N}_\mathcal{L}(\lambda)\log\frac{2}{\delta}}{M}}\right)^{2r-1}\left(\frac{4\kappa^2\beta_\infty}{3M\lambda} + \sqrt{\frac{2p\kappa^2\beta_\infty}{M\lambda}}\right)^{2-2r} \\ &\le 2\lambda^{-\frac{1}{2}}\left(\frac{4\kappa\log\frac{2}{\delta}}{\sqrt{\lambda}M} + \sqrt{\frac{4\kappa^2\mathcal{N}_\mathcal{L}(\lambda)\log\frac{2}{\delta}}{M}}\right)^{2r-1}\left(2\sqrt{\frac{2p\kappa^2\beta_\infty}{M\lambda}}\right)^{2-2r} \\ &\le \frac{C_1}{\lambda M^r} + \sqrt{\frac{C_2'\mathcal{N}_\mathcal{L}(\lambda)^{2r-1}}{M\lambda^{3-2r}}} \le \frac{C_1}{\lambda M^r} + \sqrt{\frac{C_2}{M\lambda^{3-2r+b(2r-1)}}},\end{aligned}$$

其中，在最后一个不等式中我们使用了假设 $\mathcal{N}_\mathcal{L}(\lambda) \le c_b\lambda^{-b}$，并令 $C_1 = 2(4\kappa\log\frac{2}{\delta})^{2r-1}(8p\kappa^2\beta_\infty)^{1-r}$，$C_2 = 4(4c_b\kappa^2\log^2\frac{2}{\delta})^{2r-1}(8p\kappa^2\beta_\infty)^{2-2r}$. 根据 $M \ge C_{r\lambda}^{\frac{1}{1}}$ 和 $M \ge \frac{C_2}{\lambda^{1+b(2r-1)}}$，我们得到

$$\left\|\left(\mathcal{L}_{M,\lambda}^{-1} - \mathcal{L}_\lambda^{-1}\right)\mathcal{L}_\lambda^r\right\| \le \frac{C_1}{\lambda M^r} + \sqrt{\frac{C_2}{M\lambda^{3-2r+b(2r-1)}}} \le 2\lambda^{r-1}.$$

将此界代入（B.24）可得

$$\left\|\mathcal{L}_{M,\lambda}^{-(r\vee 1)}\mathcal{L}_\lambda^r\right\| \le 3\lambda^{r-1}.$$

• 情形（$r \geq 1$）：

$$\begin{aligned}\left\|\mathcal{L}_{M,\lambda}^{-(r\vee 1)}\mathcal{L}_\lambda^r\right\| &= \left\|\mathcal{L}_{M,\lambda}^{-r}\mathcal{L}_\lambda^r\right\| \\ &\leq 1+\left\|\mathcal{L}_{M,\lambda}^{-r}\left(\mathcal{L}_\lambda^r-\mathcal{L}_{M,\lambda}^r\right)\right\| \\ &\leq 1+\lambda^{-r}C_{\kappa,r}\|\mathcal{L}_\lambda-\mathcal{L}_{M,\lambda}\|,\end{aligned}$$

其中 $C_{\kappa,\,r}$ 由命题 B.6定义。因此，根据命题 B.21 (E_6) 的界，我们得到

$$\begin{aligned}&\left\|\mathcal{L}_{M,\lambda}^{-(r\vee 1)}\mathcal{L}_\lambda^r\right\| \\ &\leq 1+\lambda^{-r}C_{1,r}\left(\frac{2\kappa^2}{M}+\frac{2\kappa^2}{\sqrt{M}}\right)\log\frac{2}{\delta}\leq 3,\end{aligned}$$

其中使用了 $M \geq C_3\lambda^{-2r}$，且 $C_3 := 4\kappa^4 C_{1,\,r}^2 \log^2\frac{2}{\delta}$。

□

命题 B.17。 对于任意 $q > 0$, $n \geq \max\{8C_{\kappa,q}^2\kappa^4\lambda^{-2q}\log^2\frac{2}{\delta}, 100\kappa^2\mathcal{N}_\mathcal{L}(\lambda)\lambda^{-1}\log^3\frac{2}{\delta}\}$ 以及 $M \geq \frac{8p\kappa^2\beta_\infty}{\lambda}$，以至少 $1-\delta$ 的概率有，

$$\left\|\widehat{\Sigma}_{M,\lambda}^{-q}\Sigma_{M,\lambda}^q\right\| \leq 2.$$

注意，我们在 (B.11) 中对方差项使用此界，其中 $q = (r\vee 1) - \frac{1}{2}$。此外，关于 n 的上述假设在情形 $q = (r\vee 1)-\frac{1}{2}$ (当 $r \geq \frac{1}{2}$) 下由命题 B.2中给出的样本量条件自动满足，即

$$n \geq C_\bullet \log^{3(2r+b)}\left(\delta^{-1}\right)\lambda^{-(2r+b)}.$$

证明。• **情形** $q < 1$**:**

由命题 B.21(E_3)，我们以至少 $1-\delta$ 的概率得到

$$\begin{aligned}\left\|\widehat{\Sigma}_{M,\lambda}^{-q}\Sigma_{M,\lambda}^q\right\| &= \left\|\widehat{\Sigma}_{M,\lambda}^{-1}\Sigma_{M,\lambda}\right\|^q \\ &\leq \left\|\widehat{\Sigma}_{M,\lambda}^{-1}\left(\widehat{\Sigma}_M-\Sigma_M\right)\right\|_{HS}+1 \\ &\leq \frac{1}{\sqrt{\lambda}}\left\|\widehat{\Sigma}_{M,\lambda}^{-\frac{1}{2}}\Sigma_{M,\lambda}^{\frac{1}{2}}\right\|\left\|\Sigma_{M,\lambda}^{-\frac{1}{2}}\left(\widehat{\Sigma}_M-\Sigma_M\right)\right\|_{HS}+1 \\ &\leq \frac{1}{\sqrt{\lambda}}\left\|\widehat{\Sigma}_{M,\lambda}^{-\frac{1}{2}}\Sigma_{M,\lambda}^{\frac{1}{2}}\right\|\left(\frac{2\kappa}{\sqrt{\lambda}n}+\sqrt{\frac{4\kappa^2\mathcal{N}_{\mathcal{L}_M}(\lambda)}{n}}\right)\log\frac{2}{\delta}+1.\end{aligned}$$

由命题 B.15，我们以至少 $1-\delta$ 的概率得到

$$\left\|\widehat{\Sigma}_{M,\lambda}^{-\frac{1}{2}}\Sigma_{M,\lambda}^{\frac{1}{2}}\right\| \leq 2$$

因此

$$\left\|\widehat{\Sigma}_{M,\lambda}^{-q}\Sigma_{M,\lambda}^q\right\| \leq \frac{2}{\sqrt{\lambda}}\left(\frac{2\kappa}{\sqrt{\lambda}n}+\sqrt{\frac{4\kappa^2\mathcal{N}_{\mathcal{L}_M}(\lambda)}{n}}\right)\log\frac{2}{\delta}+1. \tag{B.25}$$

由 B.18，我们以至少 $1-2\delta$ 的概率得到

$$\mathcal{N}_{\mathcal{L}_M}(\lambda) \leq \left(1+2\log\frac{2}{\delta}\right)4\mathcal{N}_\mathcal{L}(\lambda).$$

将此界代入（B.25）可得

$$\left\|\widehat{\Sigma}_{M,\lambda}^{-q}\Sigma_{M,\lambda}^q\right\| \leq \frac{2}{\sqrt{\lambda}}\left(\frac{2\kappa}{\sqrt{\lambda}n}+\sqrt{\frac{4\kappa^2\left(1+2\log\frac{2}{\delta}\right)4\mathcal{N}_\mathcal{L}(\lambda)}{n}}\right)\log\frac{2}{\delta}+1 \leq 2, \tag{B.26}$$

其中在最后一个不等式中，我们使用了 $n \geq 100\kappa^2\mathcal{N}_\mathcal{L}(\lambda)\lambda^{-1}\log^3\frac{2}{\delta}$。

• **情形** $q \geq 1$：由命题 B.6和命题 B.21(E_7)，我们以至少 $1-\delta$ 的概率，

$$\begin{aligned}\left\|\widehat{\Sigma}_{M,\lambda}^{-q}\Sigma_{M,\lambda}^{q}\right\| &\leq \lambda^{-q}\left\|\widehat{\Sigma}_M^q - \Sigma_M^q\right\|_{HS} + 1 \\ &\leq \lambda^{-q}C_{\kappa,q}\left\|\widehat{\Sigma}_M - \Sigma_M\right\|_{HS} + 1 \\ &\leq \lambda^{-q}C_{\kappa,q}\left(\frac{2\kappa^2}{n} + \frac{2\kappa^2}{\sqrt{n}}\right)\log\frac{2}{\delta} + 1 \leq 2,\end{aligned}$$

其中，我们在最后一个不等式中使用了$n \geq 8C_{\kappa,q}^2\kappa^4\lambda^{-2q}\log^2\frac{2}{\delta}$。

□

命题 B.18. 对任意 $M \geq \frac{8p\kappa^2\beta_\infty}{\lambda}$，我们以至少 $1-2\delta$ 的概率有，

$$\mathcal{N}_{\mathcal{L}_M}(\lambda) \leq \left(1 + 2\log\frac{2}{\delta}\right)4\mathcal{N}_{\mathcal{L}}(\lambda).$$

证明。

$$\begin{aligned}\mathcal{N}_{\mathcal{L}_M}(\lambda) &\leq \operatorname{Tr}[\mathcal{L}_M\mathcal{L}_\lambda^{-1}]\left\|\mathcal{L}_\lambda^{\frac{1}{2}}\mathcal{L}_{M,\lambda}^{-\frac{1}{2}}\right\|^2 \\ &= \left(\mathcal{N}_{\mathcal{L}} + \operatorname{Tr}[(\mathcal{L}_M - \mathcal{L})\mathcal{L}_\lambda^{-1}]\right)\left\|\mathcal{L}_\lambda^{\frac{1}{2}}\mathcal{L}_{M,\lambda}^{-\frac{1}{2}}\right\|^2 \\ &= \left(\mathcal{N}_{\mathcal{L}} + \|B\|_{HS}\right)\left\|\mathcal{L}_\lambda^{\frac{1}{2}}\mathcal{L}_{M,\lambda}^{-\frac{1}{2}}\right\|^2,\end{aligned}$$

其中 $B := \mathcal{L}_\lambda^{-\frac{1}{2}}(\mathcal{L}_M - \mathcal{L})\mathcal{L}_\lambda^{-\frac{1}{2}}$。命题 B.21（$E_4$）我们以至少 $1-\delta$ 的概率有

$$\|B\|_{HS} \leq 2\left(\frac{2\kappa^2}{\lambda M} + \sqrt{\frac{\kappa^2\mathcal{N}_{\mathcal{L}}(\lambda)}{\lambda M}}\right)\log\frac{2}{\delta}.$$

利用 $\lambda > 4\kappa^2 M^{-1}$，我们得到

$$\|B\|_{HS} \leq 2\mathcal{N}_{\mathcal{L}}(\lambda)\log\frac{2}{\delta}.$$

此外，由命题 B.14 我们以至少 $1-\delta$ 的概率有，

$$\left\|\mathcal{L}_\lambda^{\frac{1}{2}}\mathcal{L}_{M,\lambda}^{-\frac{1}{2}}\right\|^2 \leq 4.$$

综上，我们得到

$$\mathcal{N}_{\mathcal{L}_M}(\lambda) \leq \left(\mathcal{N}_{\mathcal{L}} + \|B\|_{HS}\right)\left\|\mathcal{L}_\lambda^{\frac{1}{2}}\mathcal{L}_{M,\lambda}^{-\frac{1}{2}}\right\|^2 \leq \left(1 + 2\log\frac{2}{\delta}\right)4\mathcal{N}_{\mathcal{L}}(\lambda).$$

□

B.6 集中不等式

命题 B.19. 设 $\mathcal{X}_1, \cdots, \mathcal{X}_m$ 是可分希尔伯特空间上的一列独立同分布的自伴希尔伯特–施密特算子。假设 $\mathbb{E}[\mathcal{X}_1] = 0$，且 $\|\mathcal{X}_1\| \leq B$ 几乎必然成立，其中某个 $B > 0$。设 $\mathcal{V}$ 是一个正迹类算子，满足 $\mathbb{E}\left[\mathcal{X}_1^2\right] \preccurlyeq \mathcal{V}$。那么，以至少 $1-\delta$ 的概率，$(\delta \in]0,1[)$，有

$$\left\|\frac{1}{m}\sum_{i=1}^{m}\mathcal{X}_i\right\| \leq \frac{2B\beta}{3m} + \sqrt{\frac{2\|\mathcal{V}\|\beta}{m}}, \quad \beta = \log\frac{4\operatorname{tr}\mathcal{V}}{\|\mathcal{V}\|\delta}.$$

证明。 该命题最初由 Tropp (2011针对矩阵情形证明。对于包含算子在内的一般情形，证明可见Lin 和 Cevher (2018)（见引理 26）。

□

命题 **B.20.** 以下是针对希尔伯特空间值随机变量的集中性结果，可以可在（Caponnetto 和 De Vito，2007年Caponnetto 和 De Vito （2007年））。

设 w_1、$\cdots$、w_n 是取值于可分希尔伯特空间中的独立同分布随机变量，其范数为 $\|.\|$。假设存在两个正数常数 B 和 σ^2，使得

$$\mathbb{E}\Big[\|w_1 - \mathbb{E}[w_1]\|^l\Big] \leq \frac{1}{2} l! B^{l-2} V^2, \quad \forall l \geq 2. \tag{B.27}$$

那么，对于任意 $0 < \delta < 1/2$，以下结论以至少 $1-\delta$ 的概率成立，

$$\left\| \frac{1}{n} \sum_{k=1}^{n} w_n - \mathbb{E}[w_1] \right\| \leq \left(\frac{2B}{n} + \frac{2V}{\sqrt{n}} \right) \log \frac{2}{\delta}.$$

特别地，若*B.27*成立，则

$$\|w_1\| \leq B/2 \quad a.s., \quad and \quad \mathbb{E}\Big[\|w_1\|^2\Big] \leq V^2.$$

命题 B.21. 对于任意 $\lambda > 0$，定义以下事件，

$$E_1 = \left\{ \left\| \Sigma_{M,\lambda}^{-\frac{1}{2}} \left(\widehat{\Sigma}_M - \Sigma_M \right) \Sigma_{M,\lambda}^{-\frac{1}{2}} \right\| \leq \frac{4\kappa^2 \beta_M}{3n\lambda} + \sqrt{\frac{2\kappa^2 \beta_M}{n\lambda}} \right\}, \qquad \beta_M = \log \frac{4\kappa^2 (\mathcal{N}_{\mathcal{L}_M}(\lambda) + 1)}{\delta \|\mathcal{L}_M\|},$$

$$E_2 = \left\{ \left\| \mathcal{L}_\lambda^{-\frac{1}{2}} (\mathcal{L}_M - \mathcal{L}) \mathcal{L}_\lambda^{-\frac{1}{2}} \right\| \leq \frac{4\kappa^2 \beta_\infty}{3M\lambda} + \sqrt{\frac{2p\kappa^2 \beta_\infty}{M\lambda}} \right\}, \qquad \beta_\infty = \log \frac{4\kappa^2 (\mathcal{N}_{\mathcal{L}}(\lambda) + 1)}{\delta \|\mathcal{L}\|},$$

$$E_3 = \left\{ \left\| \Sigma_{M,\lambda}^{-\frac{1}{2}} \left(\widehat{\Sigma}_M - \Sigma_M \right) \right\|_{HS} \leq \left(\frac{2\kappa}{\sqrt{\lambda} n} + \sqrt{\frac{4\kappa^2 \mathcal{N}_{\mathcal{L}_M}(\lambda)}{n}} \right) \log \frac{2}{\delta} \right\},$$

$$E_4 = \left\{ \left\| \mathcal{L}_\lambda^{-\frac{1}{2}} (\mathcal{L}_M - \mathcal{L}) \mathcal{L}_\lambda^{-\frac{1}{2}} \right\|_{HS} \leq \left(\frac{4\kappa^2}{\lambda M} + \sqrt{\frac{4\kappa^2 \mathcal{N}_{\mathcal{L}}(\lambda)}{\lambda M}} \right) \log \frac{2}{\delta} \right\},$$

$$E_5 = \left\{ \left\| \mathcal{L}_\lambda^{-\frac{1}{2}} (\mathcal{L}_M - \mathcal{L}) \right\| \leq \left(\frac{2\kappa}{\sqrt{\lambda} M} + \sqrt{\frac{4\kappa^2 \mathcal{N}_{\mathcal{L}}(\lambda)}{M}} \right) \log \frac{2}{\delta} \right\},$$

$$E_6 = \left\{ \|\mathcal{L} - \mathcal{L}_M\|_{HS} \leq \left(\frac{2\kappa^2}{M} + \frac{2\kappa^2}{\sqrt{M}} \right) \log \frac{2}{\delta} \right\},$$

$$E_7 = \left\{ \left\| \widehat{\Sigma}_M - \Sigma_M \right\|_{HS} \leq \left(\frac{2\kappa^2}{n} + \frac{2\kappa^2}{\sqrt{n}} \right) \log \frac{2}{\delta} \right\}.$$

在假设 *3.1* 成立的条件下，对于任意 $\delta \in (0,1)$，上述每个事件均以至少 $1-\delta$ 的概率成立。

证明。关于 E_1 的界可完全按照 Lin 和 Cevher (2018)（引理 18）的证明步骤得到。事件 $E_2 - E_7$ 的界已由 Rudi 和 Rosasco (2016年)（见命题 6、引理 8 和命题 10）给出。然而，由于假设和设定不同，我们得到的界略有不同，因此为完整起见，给出事件 $E_2 - E_7$ 的证明。

E_2) 首先注意到 $\mathcal{L}_M$ 可表示为

$$\mathcal{L}_M = \frac{1}{M} \sum_{m=1}^{M} \sum_{i=1}^{p} \varphi_m^{(i)} \otimes \varphi_m^{(i)},$$

其中 $\varphi_m(\cdot) = \varphi(\cdot, \omega_m)$，且张量积是相对于 $L^2(\mathcal{U}, \rho_\mathcal{U})$ 内积取的。

上述恒等式可由直接计算得出:

$$
\begin{aligned}
(\mathcal{L}_M G)(u) &= \int K_M(u,\tilde{u})\, G(\tilde{u})\, d\rho_{\mathcal{U}}(\tilde{u}) \\
&= \frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{p}\int \varphi_m^{(i)}(u) \otimes_{\mathcal{V}} \varphi_m^{(i)}(\tilde{u})\, G(\tilde{u})\, d\rho_{\mathcal{U}}(\tilde{u}) \\
&= \frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{p}\varphi_m^{(i)}(u)\int \big\langle \varphi_m^{(i)}(\tilde{u}), G(\tilde{u})\big\rangle_{\mathcal{V}} d\rho_{\mathcal{U}}(\tilde{u}) \\
&= \frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{p}\big(\varphi_m^{(i)} \otimes \varphi_m^{(i)}\big)(G)(u).
\end{aligned}
$$

类似地，我们有 $\mathcal{L} = \mathrm{E}[\sum_{i=1}^{p} \varphi_i \otimes \varphi_i]$。
现在定义 $\mathcal{X}_m := \mathcal{L}_\lambda^{-\frac{1}{2}}(\mathcal{L}_M^{(m)} - \mathcal{L})\mathcal{L}_\lambda^{-\frac{1}{2}}$，其中 $\mathcal{L}_M^{(m)} := \sum_{i=1}^{p} \varphi_m^{(i)} \otimes \varphi_m^{(i)}$。我们现在得到

$$
\|\mathcal{X}_1\| \le \left\|\mathcal{L}_\lambda^{-\frac{1}{2}}\mathcal{L}_M^{(m)}\mathcal{L}_\lambda^{-\frac{1}{2}}\right\| + \mathbb{E}\left\|\mathcal{L}_\lambda^{-\frac{1}{2}}\mathcal{L}_M^{(m)}\mathcal{L}_\lambda^{-\frac{1}{2}}\right\| \le 2\frac{\kappa^2}{\lambda} := B,
$$

其中，对于最后一个不等式，我们使用了

$$
\left\|\mathcal{L}_\lambda^{-\frac{1}{2}}\mathcal{L}_M^{(m)}\mathcal{L}_\lambda^{-\frac{1}{2}}\right\| \le \lambda^{-1}\left\|\mathcal{L}_M^{(m)}\right\| \le \frac{\kappa^2}{\lambda}.
$$

对于二阶矩，由詹森不等式我们有

$$
\begin{aligned}
\mathbb{E}\big[\mathcal{X}^2\big] &\preccurlyeq \mathbb{E}\left[\left(\mathcal{L}_\lambda^{-\frac{1}{2}}\mathcal{L}_M^{(m)}\mathcal{L}_\lambda^{-\frac{1}{2}}\right)^2\right] \\
&\preccurlyeq \mathbb{E}\left[p\sum_{i=1}^{p}\left(\mathcal{L}_\lambda^{-\frac{1}{2}}\varphi_m^{(i)} \otimes \varphi_m^{(i)}\mathcal{L}_\lambda^{-\frac{1}{2}}\right)^2\right] \\
&= \mathbb{E}\left[p\sum_{i=1}^{p}\left\|\mathcal{L}_\lambda^{-\frac{1}{2}}\varphi_m^{(i)}\right\|_{L^2_{\rho_x}}^2 \mathcal{L}_\lambda^{-\frac{1}{2}}\varphi_m^{(i)} \otimes \varphi_m^{(i)}\mathcal{L}_\lambda^{-\frac{1}{2}}\right] \\
&\preccurlyeq \mathbb{E}\left[p\frac{\kappa^2}{\lambda}\mathcal{L}_\lambda^{-\frac{1}{2}}\mathcal{L}_M^{(m)}\mathcal{L}_\lambda^{-\frac{1}{2}}\right] \\
&= \frac{p\kappa^2}{\lambda}\mathcal{L}\mathcal{L}_\lambda^{-1} := \mathcal{V}
\end{aligned}
$$

对于 $\beta = \log\frac{4\mathcal{N}_{\mathcal{L}}(\lambda)}{\|\mathcal{V}\|\delta}$，我们有

$$
\begin{aligned}
\beta &= \log\frac{4\mathcal{N}_{\mathcal{L}}(\lambda)}{\|\mathcal{L}\mathcal{L}_\lambda^{-1}\|\delta} \\
&= \log\frac{4\mathcal{N}_{\mathcal{L}}(\lambda)(\|\mathcal{L}\| + \lambda)}{\|\mathcal{L}\|\delta} \\
&\le \log\frac{4\mathcal{N}_{\mathcal{L}}(\lambda)\|\mathcal{L}\| + 4\operatorname{tr}\mathcal{L}}{\|\mathcal{L}\|\delta} \le \log\frac{4\kappa^2(\mathcal{N}_{\mathcal{L}}(\lambda)+1)}{\|\mathcal{L}\|\delta}.
\end{aligned}
$$

该结论现在由命题B.19得出。

$_1E_3$) 设定 $w_i := \Sigma_{M,\lambda}^{-2}\xi_i$，其中 $\xi_i := K_{M,u_i}K_{M,u_i}^*$。注意 $\mathrm{E}[\xi_i] = \Sigma_M$。于是，我们得到

$$
\begin{aligned}
\|w_i\|_{\mathrm{HS}} &= \left\|\Sigma_{M,\lambda}^{-\frac{1}{2}}K_{M,u_i}K_{M,u_i}^*\right\|_{\mathrm{HS}} \\
&\le \|\Sigma_{M,\lambda}^{-\frac{1}{2}}\|\,\|K_{M,u_i}K_{M,u_i}^*\|_{\mathrm{HS}} \\
&\le \lambda^{-\frac{1}{2}}\,\|K_M(u_i,u_i)\|_{\mathrm{HS}} \;\le\; \frac{\kappa^2}{\sqrt{\lambda}} \;=: B.
\end{aligned}
$$

对于二阶矩，我们得到

$$\begin{aligned}\mathbb{E}\|w_i\|_{\mathrm{HS}}^2 &= \mathbb{E}\ \mathrm{tr}\Big[\Sigma_{M,\lambda}^{-\frac{1}{2}}\xi_i\Sigma_{M,\lambda}^{-1}\xi_i\Sigma_{M,\lambda}^{-\frac{1}{2}}\Big]\\ &\le \kappa^2\,\mathbb{E}\ \mathrm{tr}\Big[\Sigma_{M,\lambda}^{-\frac{1}{2}}K_{M,u_i}K_{M,u_i}^*\Sigma_{M,\lambda}^{-\frac{1}{2}}\Big]\\ &= \kappa^2\ \mathrm{tr}\big(\Sigma_{M,\lambda}^{-1}\Sigma_M\big) = \kappa^2\,\mathcal{N}_{\mathcal{L}_M}(\lambda)\ =: V^2.\end{aligned}$$

由此可直接应用命题B.20得出该结论。
E_4) 令 $w_m := \mathcal{L}_\lambda^{-\frac{1}{2}}(\mathcal{L}_M^{(m)} - \mathcal{L})\mathcal{L}_\lambda^{-\frac{1}{2}}$。注意，我们有

$$\begin{aligned}\|w_m\|_{HS} &\le \left\|\mathcal{L}_\lambda^{-\frac{1}{2}}\mathcal{L}_M^{(m)}\mathcal{L}_\lambda^{-\frac{1}{2}}\right\|_{HS} + \mathrm{tr}\big[\mathcal{L}\mathcal{L}_\lambda^{-1}\big]\\ &\le \left\|\mathcal{L}_\lambda^{-\frac{1}{2}}\left(\sum_{i=1}^p \varphi_m^{(i)}\otimes\varphi_m^{(i)}\right)\mathcal{L}_\lambda^{-\frac{1}{2}}\right\|_{HS} + \mathcal{N}_{\mathcal{L}}(\lambda)\\ &\le \lambda^{-1}\sum_{i=1}^p\left\|\varphi_m^{(i)}\otimes\varphi_m^{(i)}\right\|_{HS} + \mathcal{N}_{\mathcal{L}}(\lambda)\\ &\le \lambda^{-1}\sum_{i=1}^p\left\|\varphi_m^{(i)}\right\|_{L^2_{\rho_x}}^2 + \mathcal{N}_{\mathcal{L}}(\lambda) \le \frac{2\kappa^2}{\lambda} =: B.\end{aligned}$$

对于二阶矩，我们有，

$$\mathbb{E}\|w_m\|_{HS}^2 \le \mathbb{E}\,\mathrm{tr}\left[\left(\mathcal{L}_\lambda^{-\frac{1}{2}}\mathcal{L}_M^{(m)}\mathcal{L}_\lambda^{-\frac{1}{2}}\right)^2\right] \le \frac{\kappa^2}{\lambda}\mathbb{E}\,\mathrm{tr}\left[\mathcal{L}_\lambda^{-\frac{1}{2}}\mathcal{L}_M^{(m)}\mathcal{L}_\lambda^{-\frac{1}{2}}\right] = \frac{\kappa^2}{\lambda}\mathcal{N}_{\mathcal{L}}(\lambda) =: V^2,$$

其中最后一个不等式用到了 $\|\mathcal{L}_\lambda^{-\frac{1}{2}}\mathcal{L}_M^{(m)}\mathcal{L}_\lambda^{-\frac{1}{2}}\| \le \frac{\kappa^2}{\lambda}$。现在由命题B.20即可得出该结论。$E_5$) 令 $w_m := \mathcal{L}_\lambda^{-\frac{1}{2}}\mathcal{L}_M^{(m)}$。注意，我们有

$$\begin{aligned}\|w_m\|_{HS} &\le \left\|\mathcal{L}_\lambda^{-\frac{1}{2}}\mathcal{L}_M^{(m)}\right\|_{HS}\\ &\le \left\|\mathcal{L}_\lambda^{-\frac{1}{2}}\left(\sum_{i=1}^p \varphi_m^{(i)}\otimes\varphi_m^{(i)}\right)\right\|_{HS}\\ &\le \lambda^{-1/2}\sum_{i=1}^p\left\|\varphi_m^{(i)}\right\|_{L^2_{\rho_x}}^2 \le \frac{\kappa^2}{\sqrt{\lambda}} =: B.\end{aligned}$$

对于二阶矩，我们有，

$$\mathbb{E}\|w_m\|_{HS}^2 \le \kappa^2\mathbb{E}\|\mathcal{L}_\lambda^{-\frac{1}{2}}\mathcal{L}_M^{(m)}\mathcal{L}_\lambda^{-\frac{1}{2}}\|_{HS} \le \kappa^2\mathbb{E}\,\mathrm{tr}\left[\mathcal{L}_\lambda^{-\frac{1}{2}}\mathcal{L}_M^{(m)}\mathcal{L}_\lambda^{-\frac{1}{2}}\right] = \kappa^2\mathcal{N}_{\mathcal{L}}(\lambda) =: V^2$$

结合算子范数可由希尔伯特-施密特范数界定这一事实，即可由命题 B.20得出该结论： $\|.\| \le \|.\|_{HS}$。

E_6) 设 $w_m := \mathcal{L}_M^{(m)}$。注意，我们有

$$\begin{aligned}\|w_m\|_{HS} &\le \left\|\mathcal{L}_M^{(m)}\right\|_{HS} = \left\|\sum_{i=1}^p \varphi_m^{(i)}\otimes\varphi_m^{(i)}\right\|_{HS}\\ &\le \sum_{i=1}^p\left\|\varphi_m^{(i)}\right\|_{L^2_{\rho_x}}^2 \le \kappa^2 =: B.\end{aligned}$$

对于二阶矩，我们有，

$$\mathbb{E}\|w_m\|_{HS}^2 \le \kappa^4 =: V^2.$$

现在由命题 B.20E_7) 令 $w_i := \xi_i = K_{M,}$
$_{x_i}K^*_{M,\ x_i\circ}$ 注意,

$$\|w_i\|_{HS} = \left\|K_{M,x_i}K^*_{M,x_i}\right\|_{HS} \le \kappa^2 =: B.$$

对于二阶矩，我们有,

$$\mathbb{E}\|w_i\|^2_{HS} \le \kappa^4 =: V^2.$$

现在由命题B.20即可得出该结论。 □

命题 22。 若满足假设 *3.1*，则以下事件以至少 $1-\delta$:

$$E_8 = \left\{\left\|\Sigma^{-\frac{1}{2}}_{M,\lambda}\widehat{\mathcal{S}}^*_M(\mathbf{v}-\bar{G}_\rho)\right\|_{\mathcal{H}_M} \le \left(\frac{4QZ\kappa}{\sqrt{\lambda}n} + \frac{4Q\sqrt{\mathcal{N}_{\mathcal{L}_M}(\lambda)}}{\sqrt{n}}\right)\log\frac{2}{\delta}\right\}.$$

证明。 我们使用命题B.20来证明该结论。定义

$$w_i := \Sigma^{-\frac{1}{2}}_{M,\lambda}K_{M,u_i}\big(v_i - G_\rho(u_i)\big).$$

注意，$\mathrm{E}[w_i] = 0$ 和

$$\frac{1}{n}\sum_{i=1}^n w_i = \Sigma^{-\frac{1}{2}}_{M,\lambda}\widehat{\mathcal{S}}^*_M(\mathbf{v}-\bar{G}_\rho).$$

此外，根据假设 3.1，我们有

$$\begin{aligned}
\mathbb{E}\big[\|w\|^l_{\mathcal{H}_M}\big] &= \int_{\mathcal{U}}\int_{\mathcal{V}} \|v - G_\rho(u)\|^l_{\mathcal{V}}\,\rho(dv|u)\,\|\Sigma^{-\frac{1}{2}}_{M,\lambda}K_{M,u}\|^l\rho_{\mathcal{U}}(du)\\
&\le 2^{l-1}\int_{\mathcal{U}}\int_{\mathcal{V}}\big(\|v\|^l_{\mathcal{V}} + Q^l\big)\rho(dv|u)\,\|\Sigma^{-\frac{1}{2}}_{M,\lambda}K_{M,u}\|^l\rho_{\mathcal{U}}(du)\\
&\le 2^{l-1}\big(\tfrac{1}{2}l!Z^{l-2}Q^2 + Q^l\big)\int_{\mathcal{U}}\|\Sigma^{-\frac{1}{2}}_{M,\lambda}K_{M,u}\|^l\rho_{\mathcal{U}}(du)\\
&\le 2^{l-1}\big(\tfrac{1}{2}l!Z^{l-2}Q^2 + Q^l\big)\sup_{u\in\mathcal{U}}\|\Sigma^{-\frac{1}{2}}_{M,\lambda}K_{M,u}\|^{l-2}\int_{\mathcal{U}}\mathrm{tr}\big(\Sigma^{-1}_{M,\lambda}K_{M,u}K^*_{M,u}\big)\rho_{\mathcal{U}}(du)\\
&\le 2^{l-1}\big(\tfrac{1}{2}l!Z^{l-2}Q^2 + Q^l\big)\Big(\tfrac{\kappa}{\sqrt{\lambda}}\Big)^{l-2}\mathrm{tr}\Big(\Sigma^{-1}_{M,\lambda}\int_{\mathcal{U}}K_{M,u}K^*_{M,u}\rho_{\mathcal{U}}(du)\Big)\\
&\le \tfrac{1}{2}l!\Big(\tfrac{2QZ\kappa}{\sqrt{\lambda}}\Big)^{l-2}\Big(2Q\sqrt{\mathcal{N}_{\mathcal{L}_M}(\lambda)}\Big)^2\\
&= \tfrac{1}{2}l!\,B^{l-2}V^2.
\end{aligned}$$

因此，该结论可直接由命题 B.20 得出。 □

命题 B.23。 设 $\|G_\rho\|_\infty \le Q$ ，且命题 *B.11* 中的界成立:

$$\|F^*_\lambda\|_\infty \le C_{\kappa,R,D}\,\lambda^{-(\frac{1}{2}-r)^+}, \quad \textit{where } C_{\kappa,R,D} = 2\kappa^{2r+1}RD.$$

那么，下列事件以至少 $1-\delta$ 的概率成立:

$$E_9 = \left\{\left|\frac{1}{n}\|\bar{G}_\rho - \widehat{\mathcal{S}}_M F^*_\lambda\|^2_2 - \|G_\rho - \mathcal{S}_M F^*_\lambda\|^2_{L^2(\rho_{\mathcal{U}})}\right| \le 2\Big(\frac{B_\lambda}{n} + \frac{V_\lambda}{\sqrt{n}}\Big)\log\frac{2}{\delta}\right\},$$

其中

$$B_\lambda := 4\Big(Q^2 + C^2_{\kappa,R,D}\,\lambda^{-2(\frac{1}{2}-r)^+}\Big), \qquad V_\lambda := \sqrt{2}\Big(Q + C_{\kappa,R,D}\,\lambda^{-(\frac{1}{2}-r)^+}\Big)\|G_\rho - \mathcal{S}_M F^*_\lambda\|_{L^2(\rho_{\mathcal{U}})}.$$

证明。 我们应用命题B.20。定义

$$w_i := \big\|G_\rho(u_i) - F^*_\lambda(u_i)\big\|^2_{\mathcal{V}}.$$

则$\mathbb{E}[w_i] = \|G_\rho - \mathcal{S}_M F^*_\lambda\|^2_{L^2(\rho_{\mathcal{U}})}$，因此

$$\left|\frac{1}{n}\sum_{i=1}^{n} w_i - \mathbb{E}[w_1]\right| = \left|\frac{1}{n}\big\|\bar{G}_\rho - \widehat{\mathcal{S}}_M F^*_\lambda\big\|_2^2 - \big\|G_\rho - \mathcal{S}_M F^*_\lambda\big\|^2_{L^2(\rho_{\mathcal{U}})}\right|.$$

接下来只需界定 $|w_i|$ 与 $\mathbb{E}[w_1^2]$。利用 $\|G_\rho\|_\infty := \sup_{u\in\mathcal{U}} \|G_\rho(u)\|_{\mathcal{V}} \le Q$ 以及命题 B.11，我们得到

$$|w_i| \le 2\Big(Q^2 + C^2_{\kappa,R,D}\,\lambda^{-2(\frac{1}{2}-r)^+}\Big),$$

并且进一步，

$$\begin{aligned}\mathbb{E}\big[w_1^2\big] &\le 2\Big(Q^2 + C^2_{\kappa,R,D}\,\lambda^{-2(\frac{1}{2}-r)^+}\Big)\mathbb{E}[w_1] \\ &= 2\Big(Q^2 + C^2_{\kappa,R,D}\,\lambda^{-2(\frac{1}{2}-r)^+}\Big)\big\|G_\rho - \mathcal{S}_M F^*_\lambda\big\|^2_{L^2(\rho_{\mathcal{U}})}.\end{aligned}$$

因此，该结论直接得自命题B.20。 □