arXiv:2302.03459v2 [cs.LG] 1 Mar 2023

论多元样条与无限宽神经网络之间的关系

Francis Bach Inria，巴黎高等
师范学院 PSL Research
Universityfrancis.bach@inria.fr

2023年3月2日

摘要

我们研究多元样条，并证明它们可以表示为随机特征展开，其形式为具有单隐藏层和齐次激活函数的无限宽神经网络，其中激活函数是修正线性单元的幂。我们证明，与之相关的函数空间是欧几里得球上的索伯列夫空间，并给出了导数范数的显式界限。这种联系为多元样条提供了一种新的随机特征展开，使得高效算法成为可能。无论在理论上还是实践中，这种随机特征展开在数值上的表现都优于通常的随机傅里叶特征。具体而言，在一维情形下，我们比较了两种随机展开所对应的杠杆分数，并证明神经网络展开具有更优的尺度变化。

1 引言

多变量非参数回归有多种方法可供选择：决策树、局部平均方法（如纳达拉亚-沃森估计或 k 近邻回归）、神经网络，以及基于正定核的方法，如平滑样条、克里金法和核岭回归（例如参见 [1, 2, 3]）。

本文基于基于核的方法与无限宽神经网络之间已知的以下关系 [4, 5]。我们考虑一个激活函数 $\sigma: \mathbb{R} \to \mathbb{R}$，以及一个定义在 $\mathbb{R}^d$ 上、形式如下的单隐藏层神经网络模型

$$f(x) = \sum_{j=1}^{m} \eta_j \sigma(w_j^\top x + b_j),$$

其中 $\eta_j \in \mathbb{R}$、$(w_j, b_j) \in \mathbb{R}^{d+1}$，对于 $j = 1, \ldots\ldots, m$。在用于拟合模型的目标函数中加入 ℓ_2 正则化，相当于采用基于正定核的核方法

$$\hat{k}(x, y) = \frac{1}{m} \sum_{j=1}^{m} \sigma(w_j^\top x + b_j)\sigma(w_j^\top y + b_j). \tag{1}$$

例如，参见 [6]，了解核方法的入门介绍。如果 m 输入权重 $(w_j,\ b_j) \in \mathbb{R}^{d+1}$ 是独立同分布抽样的，当 m 趋于无穷时，根据大数定律，k（x，y）趋于等价核

$$k(x,y) = \mathbb{E}_{(w,b)}\big[\sigma(w^\top x + b)\sigma(w^\top y + b)\big]. \tag{2}$$

无限宽神经网络与核方法之间的这种等价性已经以多种方式得到应用：

- 给定一个已知的核 k，它可以像公式（2）中那样表示为期望，我们可以使用近似核$\hat{}\ k$，如公式（1）所示，并利用其显式随机特征推导出高效算法 [5, 7]：对于 n 个观测值，我们可以通过为每个观测值计算 $n \times n$ 维特征向量来避免计算 m核矩阵，这在 n 个观测值的情况下非常有利 $m < n$。

- 给定一个已知的神经网络架构，它们可以研究使用过参数化模型的正则化性质，也就是说，隐藏神经元的数量趋于无穷 [8]。

在本文中，我们作出以下贡献，这些贡献涉及上述将核与神经网络联系起来的两种方式：

- 我们考虑具有与 $\|x-y\|_2^{2\alpha+1}$成比例的核的多元样条[9, 10]，，其中 $\alpha \in \mathbb{N}$（$\|\cdot\|_2$表示标准欧几里得范数），并证明它们具有随机特征展开，可表示为具有单隐藏层和齐次激活函数的无限宽神经网络，其中该激活函数是“修正线性单元” [11]的 α次幂。这扩展了 [8], 的早期工作，该工作证明了 $\alpha = 0$（阶跃激活函数）情况下的这种联系。

- 我们证明，相关函数空间是阶数为 $s = \frac{d+1}{2} + \alpha$ 的索伯列夫空间，且范数之间存在明确的依赖关系。

- 这一联系为多元样条提供了一种新的随机特征展开，从而能够采用高效算法。无论在理论上还是实践中，这种随机特征展开的数值表现都优于通常的随机傅里叶特征。特别地，在一维情形下，我们比较相关的杠杆分数 [12]，以对比这两种随机展开。对于所有 Mat´ern 核，这也为随机傅里叶特征展开提供了一种更高效的替代方案 [23, 页 84]。

2 从单隐藏层神经网络到正定核

我们考虑中心为 0、半径为 R 的欧几里得球 $\mathbb{B}^d(R)$，它位于 $\mathbb{R}^d$ 中，其中 $R > 0$；并考虑在 $\mathbb{B}^d(R)$ 上对实值函数进行估计。

对于 $\alpha \in \mathbb{N}$，我们考虑形式为 $\sigma(u) = (u_+)^\alpha = \max\{u, 0\}^\alpha$ 的激活函数 σ，即当 $u \leqslant 0$ 时，$\sigma(u) = 0$；当 $u > 0$ 时，$\sigma(u) = u^\alpha$，并采用通常的约定：若 $u > 0$，则 $u^0 = 1$，重点关注 $\alpha \in \{0, 1\}$。对于 $\alpha = 0$，我们得到阶跃函数 $\sigma(u) = 1_{u>0}$；对于 $\alpha = 1$，则得到修正线性单元$\sigma(u) = u_+$。

我们考虑随机分布的权重 $(w, b) \in \mathbb{R}^{d+1}$，以及正定核

$$k(x,y) = \mathbb{E}_{(w,b)}\big[\sigma(w^\top x + b)\sigma(w^\top y + b)\big].$$

由于我们使用齐次激活函数，可以对权重（w, b）进行归一化，使其具有紧支撑。可以采用多种归一化方式和分布来获得闭式公式。

$\mathbb{R}^{d+1}$中的完全球对称性。第一种归一化是写作

$$w^\top x + b = \begin{pmatrix} w \\ b/R \end{pmatrix}^\top \begin{pmatrix} x \\ R \end{pmatrix},$$

并令 $\binom{w}{b/R}$ 具有旋转不变性，例如，在维度为 $\mathbb{R}^{d+1}$的单位 ℓ_2球面上均匀分布。这会得到相应核[13, 14]的闭式公式 $\tilde{k}_d^{(\alpha)}$，其中余弦 $\varphi = \dfrac{x^\top y + R^2}{(\|x\|_2^2 + R^2)^{1/2}(\|y\|_2^2 + R^2)^{1/2}} \geqslant 0$（因此 $\varphi \in [0, \pi/2]$），当 α取较小值时，得到：

$$\begin{aligned}
\tilde{k}_d^{(0)}(x,y) &= \frac{1}{2\pi}(\pi - \varphi) \\
\tilde{k}_d^{(1)}(x,y) &= \frac{1}{2\pi(d+1)}(\|x\|_2^2 + R^2)^{1/2}(\|y\|_2^2 + R^2)^{1/2} \times \big[\sin\varphi + (\pi - \varphi)\cos\varphi\big] \\
\tilde{k}_d^{(2)}(x,y) &= \frac{1}{2\pi(d+1)(d+3)}(\|x\|_2^2 + R^2)(\|y\|_2^2 + R^2) \times \big[3\sin\varphi\cos\varphi + (\pi - \varphi)(1 + 2\cos^2\varphi)\big].
\end{aligned}$$

更一般地（见[13]），我们有：

$$\tilde{k}_d^{(\alpha)}(x,y) = \frac{1}{2\pi(d+1)(d+3)\cdots(d+2\alpha-1)}(\|x\|_2^2 + R^2)^{\alpha/2}(\|y\|_2^2 + R^2)^{\alpha/2} \times J_\alpha(\varphi),$$

其中 $J_\alpha(\varphi) = (-1)^\alpha(\text{正弦 } \varphi)^{2\alpha+1}\left(\frac{1}{\text{正弦 } \varphi}\frac{d}{d\varphi}\right)^\alpha\left(\frac{\pi-\varphi}{\text{正弦 } \varphi}\right)$，其形式为 $P_\alpha(\text{余弦 } \varphi, \text{正弦 } \varphi) + Q_\alpha(\text{余弦 } \varphi, \text{正弦 } \varphi)(\pi - \varphi)$，其中 P_α 和 Q_α 是次数小于 α的多项式。

正则化性质。相关函数空间可以通过环境维度为 $d+1$ 时的球谐函数来描述，例如，[15, 14], 通过在超球面上使用拉普拉斯算子给出了这种描述。然而，这需要对球谐函数有深入了解，而且不容易将其与 $\mathbb{R}^d$ 中经典的导数概念联系起来。请注意，也可以使用埃尔米特多项式 [16]。总的来说，我们得到定义在 $\mathbb{B}^d(R)$ 上、阶数为 $s = \frac{d+1}{2} + \alpha$ 的索伯列夫空间，但无法用导数给出显式表达式。

本文考虑另一种归一化方式，它更易于解释，也更容易与统计学文献中已有的核建立联系。该方法仅使用 $w \in \mathbb{R}^d$ 上的球对称性。

部分球对称性（在 $\mathbb{R}^d$ 上）。我们也可以选择让 $\binom{w}{b/R}$ 在笛卡尔积 $\mathbb{S}^{d-1} \times [-1, 1]$ 上均匀分布，其中 $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ 是单位 ℓ_2 球面；这种做法沿用了 [8] 为 $\alpha = 0$ 引入的归一化方式。这对应于 w 在球面 $\mathbb{S}^{d-1}$ 上均匀分布，且 b 在 $[-R, R]$ 上均匀分布。

本文的主要目标是给出该核的闭式公式，并研究其正则化性质。我们将从一维（$d = 1$）开始，并在后续章节中推广到所有维度。

因此，我们定义正定核：

$$k_d^{(\alpha)}(x,y) = \mathbb{E}_{(w,b)}\big[(w^\top x + b)_+^\alpha (w^\top y + b)_+^\alpha\big],$$

其中，(w,b) 在 $\mathbb{S}^{d-1} \times [-R,R]$ 上服从均匀分布。

3 区间上的核 $[-R,R]$ $(d=1)$

我们首先考虑 $d=1$ 的核，其中 $w \in \{-1,1\}$。通过变量代换$b \to -b$，可得：

$$\begin{aligned} k_1^{(\alpha)}(x,y) &= \frac{1}{4R}\int_{-R}^{R}(x+b)_+^\alpha (y+b)_+^\alpha db + \frac{1}{4R}\int_{-R}^{R}(-x+b)_+^\alpha (-y+b)_+^\alpha db \\ &= \frac{1}{4R}\int_{-R}^{R}(x-b)_+^\alpha (y-b)_+^\alpha db + \frac{1}{4R}\int_{-R}^{R}(b-x)_+^\alpha (b-y)_+^\alpha db. \end{aligned}$$

3.1 闭式公式

我们先考虑 $\alpha=0$ 的情形，再由此推广。对于 $\alpha=0$，通过直接积分可得：

$$\begin{aligned} k_1^{(0)}(x,y) &= \frac{1}{4R}\int_{-R}^{\min\{x,y\}} db + \frac{1}{4R}\int_{\max\{x,y\}}^{R} db = \frac{1}{2} - \frac{1}{4R}\big[\max\{x,y\} - \min\{x,y\}\big] \\ &= \frac{1}{2} - \frac{1}{4R}|x-y|. \end{aligned}$$

通过更繁琐的直接计算，可得 α 的其他较小取值对应的表达式：

$$\begin{aligned} k_1^{(1)}(x,y) &= \frac{R^2}{6} + \frac{1}{2}xy + \frac{1}{24R}|x-y|^3 \\ k_1^{(2)}(x,y) &= \frac{R^4}{10} + \frac{2R^2 xy}{3} + \frac{R^2}{6}(x^2+y^2) + \frac{1}{2}x^2y^2 - \frac{1}{120R}|x-y|^5. \end{aligned}$$

这可以扩展到以下命题中的所有 α 值，该命题见附录 B.1。请注意，在一维情况下，[17] 已经揭示了三次样条与无限宽神经网络之间的联系。

命题1（$d=1$的闭式公式）设 $\alpha \in \mathbb{N}$，对于在笛卡尔积 $\{-1,1\} \times [-R,R]$上均匀分布的(w,b)，我们有：

$$k_1^{(\alpha)}(x,y) = \mathbb{E}_{(w,b)}\big[(w^\top x + b)_+^\alpha (w^\top y + b)_+^\alpha\big] = P_1^{(\alpha)}(x^2, y^2, xy) + \frac{1}{R}c_1^{(\alpha)}|x-y|^{2\alpha+1},$$

其中 $P_1^{(\alpha)}$是一个次数为 α 的多项式，使得 $k^{(\alpha)}{}_{,1}{}^{(\mathrm{pol})}(x,y) = P_{(\alpha)1}(x^2,y^2,xy)$ 是一个正定核，并且 $c_1^{(\alpha)} = (-1)^{\alpha+1}\dfrac{(\alpha!)^2\,(2\alpha+1)!}{4}$。

此外，正如附录 B.1 所示，我们有一种特殊形式的多项式核 $k_{1}^{(\alpha),\,(\mathrm{pol})}(x,y) = P_{1}^{(\alpha)}(x^2, y^2, xy)$，如下：

$$
\begin{aligned}
k_1^{(\alpha),(\mathrm{pol})}(x,y) &= \frac{1}{4R}\int_{-R}^{R}(x-b)^{\alpha}(y-b)^{\alpha}db, \quad \text{which can be expressed as} \\
k_1^{(\alpha),(\mathrm{pol})}(x,y) &= \frac{1}{2R}\sum_{i,j=0}^{\alpha} 1_{i+j \text{ even}} \cdot \binom{\alpha}{i}\binom{\alpha}{j} x^i y^j \frac{R^{2\alpha+1-i-j}}{2\alpha+1-i-j} \\
&= \frac{1}{2}\sum_{s=0}^{\alpha}\frac{R^{2\alpha-2s}}{2\alpha+1-2s}\sum_{i,j=0}^{\alpha} 1_{i+j=2s} \cdot \binom{\alpha}{i}\binom{\alpha}{j} x^i y^j. \qquad (3)
\end{aligned}
$$

注意，对于 $s=0$，其项为 $_{2}\frac{R^{2\alpha}}{2\alpha+1}();$ 而对于 $\alpha>0$，对应于 $s=1$ 的项等于$\frac{\alpha^2 R^{2(\alpha-1)}}{2\,2\alpha-1}()\ xy$。在所有情况下，计算所需时间至多为 $O(\alpha^2)$。此外，对应的特征空间涵盖所有次数小于 α 的多项式（证明见附录B.3）。在命题3中，这一结果将直接推广到维度大于一的情形。

3.2 对应范数

所有正定核都定义了一个希尔伯特空间，其中包含定义在 $\mathbb{B}^d(R)$ 上的实值函数，并赋予其特定范数。对于可表示为期望的核，该范数 $\Omega_1^{(\alpha)}$ 等于 [18, 14]:

$$
\begin{aligned}
\Omega_1^{(\alpha)}(f)^2 &= \inf_{\eta_\pm:[-R,R]\to\mathbb{R}} \frac{1}{4R}\int_{-R}^{R}\left[\eta_+(b)^2+\eta_-(b)^2\right]db \\
&\qquad \text{such that } \forall x\in[-R,R], f(x) = \frac{1}{4R}\int_{-R}^{R}\left[\eta_+(b)(x-b)_+^{\alpha}+\eta_-(b)(b-x)_+^{\alpha}\right]db,
\end{aligned}
$$

其中，下确界取遍平方可积函数 η_+ 和 η_-。

特殊情形 $\alpha=0$。对于连续可微的 f，我们可以使用并对以下两种简单表示取平均:

$$f(x) = f(-R) + \int_{-R}^{x} f'(b)(x-b)_+^0 db = f(R) - \int_{-R}^{R} f'(b)(b-x)_+^0 db,$$

得到 $f(x) = \frac{1}{2}[f(R)+f(-R)] + 2R\int_{-R}^{R} f'(b)(x-b)_+^0\frac{db}{4R} - 2R\int_{-R}^{R} f'(b)(b-x)_+^0\frac{db}{4R}$. 在 $[-R,R]$ 上等于 $1/2$ 的常函数可以表示为:

$$\int_{-R}^{R}(x-b)_+^0\frac{db}{4R} + \int_{-R}^{R}(b-x)_+^0\frac{db}{4R}.$$

因此我们可以取 : $\eta_+(b) = 2Rf'(b) + [f(R) + f(-R)]$ 和 $\eta_+(b) = -2Rf'(b) + [f(R) + f(-R)]$。

T这就得到平方范数 $\Omega_1^{(0)}(f)^2$ 小于（因为交叉项相互抵消） :

$$2R\int_{-R}^{R} f'(x)^2 dx + \left[f(-R)+f(R)\right]^2.$$

特别是，只要上面的量有定义，即 f' 平方可积，范数就是有限的。为了证明这确实是正确的范数，我们只需检验该表示是最优的；如下文所示，对于所有 α都成立（见命题 2）。因此

$$\Omega_1^{(0)}(f)^2 = 2R\int_{-R}^{R} f'(x)^2 dx + \big[f(-R) + f(R)\big]^2.$$

一般情形 $\alpha \geqslant 0$。为了得到范数，我们可以注意到，由函数 $(x-b)_+^\alpha$ 构成的连续展开恰好可由带积分余项的泰勒展开得到；这种展开适用于定义在 $[-R,R]$ 上且具有 $\alpha+1$ 阶连续导数的函数：

$$f(x) = \sum_{i=0}^{\alpha} \frac{f^{(i)}(-R)}{i!}(x+R)^i + \int_{-R}^{R} \frac{f^{(\alpha+1)}(b)}{\alpha!}(x-b)_+^\alpha db.$$

忽略边界条件后，我们看到 $\eta_+(b)$ 应与 $\frac{1}{\alpha!}f_{(\alpha+1)}(b)$ 相关，而再生核希尔伯特空间范数应包含积分 $\int_R^R f^{(\alpha+1)}(x)^2 dx$。以下命题将明确说明这一点（证明见附录 B.2）。

命题 2（ $d=1$ 的再生核希尔伯特空间范数） 与核 $k_1^{(\alpha)}$ 相关的定义在 $[-R,R]$ 上的函数的再生核希尔伯特空间范数等于：

$$\Omega_1^{(\alpha)}(f)^2 = \frac{2R}{\alpha!^2}\int_{-R}^{R} f^{(\alpha+1)}(x)^2 dx + \Theta^{(\alpha)}\big[f^{(0)}(-R), f^{(0)}(R), \ldots, f^{(\alpha)}(-R), f^{(\alpha)}(R)\big],$$

其中 $\Theta^{(\alpha)}$ 是非负二次型。

例如，对于 $\alpha=1$，我们得到：

$$\Omega_1^{(1)}(f)^2 \quad = \quad 2R\int_{-R}^{R} f''(x)^2 dx + \big[f'(R) + f'(-R)\big]^2 + \frac{3}{2R^2}\big[f(-R) + f(R) - Rf'(R) + Rf'(-R)\big]^2.$$

与经典索博列夫范数的等价性。 利用索博列夫空间的经典结果 [19], 可以证明命题 2 中的范数等价于经典的索博列夫平方范数 $R\int_{-R}^{R} f^{(\alpha+1)}(x)^2 dx + \frac{1}{R^{2\alpha+1}}\int_{-R}^{R} f(x)^2 dx$。我们将在后续章节中将这一结果推广到所有维度，并给出明确的等价关系。

4 球上的核 $\mathbb{B}^d(R)$ $(d \geqslant 1)$

现在，我们将第 3 节的结果推广到所有维度 $d \geqslant 1$。我们将得到显式的闭式公式，但再生核希尔伯特空间范数的表达会稍微不那么明确。

4.1 闭式公式

我们先从直接推广命题 1的闭式公式开始。

命题 3（ $d \geqslant 1$的闭式公式）设 $\alpha \in \mathbb{N}$，对于在笛卡尔积 $\mathbb{S}^{d-1} \times [-1,1]$ 上均匀分布的 $\binom{w}{b/R}$，我们有，

$$k_d^{(\alpha)}(x,y) = \mathbb{E}_{(w,b)}\left[(w^\top x + b)_+^\alpha (w^\top y + b)_+^\alpha\right] = P_d^{(\alpha)}(\|x\|_2^2, \|y\|_2^2, x^\top y) + \frac{1}{R} c_d^{(\alpha)} \|x-y\|_2^{2\alpha+1},$$

其中 $P_{(\alpha)\,d}$是一个次数为 α 的多项式，使得 $k^{(\alpha)}{}_{,d}{}^{(\mathrm{pol})}(x,y) = P_d^{(\alpha)}(\|x\|_2^2, \|y\|_2^2, x^\top y)$ 是一个正定核，且 $c_d^{(\alpha)} = (-1)^{\alpha+1} \frac{\alpha!^3 \Gamma(\frac{d}{2})}{4\sqrt{\pi}\,(2\alpha+1)!\,\Gamma(\frac{d}{2}+\frac{1}{2}+\alpha)}$。

证明：我们有 $k_d^{(\alpha)}(x,y) = \mathbb{E}_w\left[k_1^{(\alpha)}(w^\top x, w^\top y)\right]$，只需对在球面上均匀分布的 w 使用以下公式：$\mathbb{E}[|w^\top z|^{2\alpha+1}] = \|z\|_2^{2\alpha+1} \frac{\Gamma(1+\alpha)\Gamma(\frac{d}{2})\,\Gamma(\frac{1}{2})\Gamma(\frac{d}{2}+\frac{1}{2}+\alpha)}$（见附录 A），从而得到 $c_d^{(\alpha)}$ 的表达式。对于多项式核部分，我们使用公式 (3)，以及以下事实：对于均匀分布的 w，且 $i+j$ 为偶数时，$\mathbb{E}\left[(w^\top x)^i (w^\top y)^j\right]$ 是关于 $x^\top x$、$y^\top y$ 和 $y^\top x$ 的次数小于 $i+j$ 的多项式。 ■

与 $d=1$ 类似，核 $P_d^{(\alpha)}(\|x\|_2^2, \|y\|_2^2, x^\top y) = k^{(\alpha)}{}_{,\,d}{}^{(\mathrm{pol})}(x,y)$ 也有积分表示，如下：

$$k_d^{(\alpha),(\mathrm{pol})}(x,y) = \frac{1}{4R}\int_{-R}^{R} \mathbb{E}_w\left[(w^\top x + b)^\alpha (w^\top y + b)^\alpha\right] db.$$

注意，我们定义了一个新的正定多项式核，它是标准核 $(x,\ y) \mapsto (R + x^\top y)^\alpha$的替代方案，计算时间为 $O(d)$（其中的常数取决于 α）。如附录B.4所示，对应的空间张成了所有次数小于 α （或等于）的多项式。

对于 $\alpha \in \{0,1,2\}$，有：

$$\begin{aligned}
k_d^{(0)}(x,y) &= \frac{1}{2} - \frac{1}{4R}\frac{\Gamma(1)\Gamma(\frac{d}{2})}{\Gamma(1/2)\Gamma(\frac{d+1}{2})}\|x-y\|_2 \\
k_d^{(1)}(x,y) &= \frac{R^2}{6} + \frac{1}{2d}x^\top y + \frac{1}{24R}\frac{\Gamma(2)\Gamma(\frac{d}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{d}{2}+\frac{3}{2})}\|x-y\|_2^3 \\
k_d^{(2)}(x,y) &= \frac{R^4}{10} + \frac{1}{3d}x^\top y + \frac{R^2}{6d}(\|x\|_2^2 + \|y\|_2^2) + \frac{1}{2d(d+2)}(2(x^\top y)^2 + \|x\|_2^2\|y\|_2^2) \\
&\qquad - \frac{1}{120R}\frac{\Gamma(2)\Gamma(\frac{d}{3})}{\Gamma(\frac{1}{2})\Gamma(\frac{d}{2}+\frac{5}{2})}\|x-y\|_2^5.
\end{aligned}$$

一个简单的界。我们需要给出关联特征的一个界。对于 $\|x\|_2 \leqslant R$，有：

$$\begin{aligned}
k_d^{(\alpha}(x,x) &= \frac{1}{4R}\int_{-R}^{R} \mathbb{E}_w\left[(w^\top x + b)^2\right]\alpha db \\
&\leqslant \frac{1}{4R}\int_{-R}^{R} \mathbb{E}_w (2R)^{2\alpha} db = \frac{1}{2}(2R)^{2\alpha}.
\end{aligned} \tag{4}$$

4.2 对应范数

在维度 $d=1$下，我们可以给出相应再生核希尔伯特空间范数的显式公式，该公式依赖于带积分余项的泰勒公式。在更高维度下，这一公式将不那么显式，因此我们需要运用多元样条理论 [9, 10]。

5 与多元样条的联系

本节先回顾样条，然后明确阐述两者之间的联系。有关多元样条的更多详情，请参见 [20, 21, 18]。

5.1 多元样条回顾

我们考虑以下函数:

$$E_d^{(\alpha)}(z) = c_d^{(\alpha)} \|z\|_2^{\alpha},$$

其傅里叶变换（定义为分布，参见 [22]）为:

$$c_d^{(\alpha)}(-1)^{\alpha+1}2^{d+1+2\alpha}\pi^{d/2-1}\Gamma(\alpha+3/2)\Gamma(d/2+1/2+\alpha)\frac{1}{\|\omega\|_2^{d+1+2\alpha}} = b_d^{(\alpha)}\frac{1}{\|\omega\|_2^{d+1+2\alpha}}.$$

核 $E_d^{(\alpha)}(x-y)$ 已知是“α 阶条件正定”的 [20, 21], 也就是说，对于每个 x_1，...，$x_n \in \mathbb{R}^d$，以及 λ_1，...，λ_n，使得 $\sum_{i=1}^n \lambda_i P(x_i) = 0$ 对所有次数小于 α 的多项式 P，

$$\sum_{i,j=1}^{n} \lambda_i \lambda_j E_d^{(\alpha)}(x_i - x_j) \geqslant 0.$$

我们还知道，对于任意函数 $L:\ \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$，以下最小化问题

$$L(f(x_1), \ldots, f(x_n)) + \frac{1}{2b_d^{(\alpha)}}\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} |\hat{f}(\omega)|^2 \|\omega\|_2^{d+1+2\alpha} d\omega$$

$_n$在 $f(x) = P(x) + \sum \lambda_i E_d^{(\alpha)}(x - x_i)$ 处取得，其中 P 和 λ 通过以下最小化问题得到：$_{i=1}$

$$L(f(x_1), \ldots, f(x_n)) + \frac{1}{2}\sum_{i,j=1}^{n} \lambda_i \lambda_j E_d^{(\alpha)}(x_i - x_j)$$

关于次数小于 α 的多项式 P，以及满足以下条件的 $\lambda \in \mathbb{R}^n$：对于所有次数小于 α [9] 的多项式 Q, $\sum_{i=1}^n \lambda_i Q(x_i) = 0$ 。

When d 是奇数时，我们可以用偏导数给出一个显式表示 :

$$\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} |\hat{f}(\omega)|^2 \|\omega\|_2^{d+1+2\alpha} d\omega = \int_{\mathbb{R}^d} \|D^{\frac{d+1}{2}+\alpha} f(x)\|^2 dx,$$

其中， $D^{\frac{d+1}{2}+\alpha} f$ 是所有 $\frac{d+1}{2}+\alpha$ 阶偏导数组成的张量。

上述展开可扩展为对 $\mathbb{B}^d(R)$ 上函数的表示，使得当 $f(x) = P(x) + \int_{\mathbb{B}^d} (R) E_d^{(\alpha)}(x-y) d\lambda(y)$ 时，上述范数有限，其中 λ 是拉东测度。

算法。给定一个映射 $\varphi\colon \mathbb{R}^d \to \mathbb{R}^m$，它能够表示所有次数小于 α 的多项式，即满足 $m = \binom{d+\alpha}{\alpha}$，那么我们可以将定义为 $y_i = f(x_i)$ 的向量 $y \in \mathbb{R}^n$ 写成 $y = \Phi\nu + K\lambda$，其中 $\Phi \in \mathbb{R}^{n\times m}$ 是包含所有 $\varphi(x_i)$、$i = 1$、.....、n 的矩阵，而 $K \in \mathbb{R}^{n\times n}$ 是与 $E_d^{(\alpha)}$ 相关联的核矩阵。我们添加约束 $\Phi^\top \lambda = 0$。条件正定意味着，当 ρ 足够大时， $K + \rho\Phi\Phi^\top$ 是半正定的。本文给出了一个显式的 $\rho\Phi\Phi^\top$，使得当数据受限于 $\mathbb{B}^d(R)$ 时这一性质成立。注意，当 $\Phi^\top\lambda = 0$ 时，我们的核的多项式部分无关紧要。

上述算法需要求解一个维度为 n 的优化问题，而我们将在下文看到，如何利用随机特征来降低其复杂度。

5.2 与索伯列夫空间的等价性

如命题3所示，我们的核等于

$$k_d^{(\alpha)}(x,y) = k_d^{(\alpha),(\mathrm{pol})}(x,y) + c_d^{(\alpha)} \|x-y\|_2^{2\alpha+1}.$$

我们将证明 再生核希尔伯特空间范数等价于 $\Omega_d^{(\alpha)(\mathrm{eq})}(f)^2$，其定义为最小值 o f

$$\frac{1}{R}\frac{1}{b_d^{(\alpha)}(2\pi)^d}\int_{\mathbb{R}^d} \|\omega\|_2^{d+1+2\alpha} |\hat{g}(\omega)|^2 d\omega + \frac{\underline{c}^2}{R^{2\alpha}} \int_{\mathbb{B}^d(R)} |g(x)|^2 dx, \tag{5}$$

对所有满足 f 在 $\mathbb{B}^d(R)$ 上等于 $\mathbb{R}^d \to \mathbb{R}$ 的函数 g 取值，其中常数 $\underline{c}$ 经过适当选取。

范数比较。我们考虑两个正的常数 $\underline{c}$ 和 $\overline{c}$，使得 :

- 对于任意次数小于 α的多项式 P，我们有 $\Omega_{(\alpha)\ d}(P) \leqslant \frac{c}{R^\alpha}\|P\|_{L_2(\mathbb{B}^d(R))}$，其中 $\underline{c}$为正常数。这样的常数必然存在，因为在有限维空间上定义范数的两个核必须具有等价的范数。目前，我们还没有得到该常数 $\underline{c}$的显式上界。
- 对于由 $k_{(\alpha)\ d}$ 定义的再生核希尔伯特空间中的任意 f， $\|f\|_{L_2(\mathbb{B}^d(R))} \leqslant R^\alpha \overline{c}\Omega_{(\alpha)\ d}(f)$。之所以必然存在这一项，是因为对于任意 $x \in \mathbb{B}^d(R)$，与所有再生核希尔伯特空间一样，我们有 $f(x)^2 \leqslant k_{(\alpha)\ d}(x,x) \cdot \Omega_{(\alpha)\ d}(f)^2$。因此，通过积分，根据公式（4），我们有 $\overline{c}^2 R^{2\alpha} \leqslant \mathrm{vol}(\mathbb{B}^d(R)) \sup_{x\in\mathbb{B}^d(R)} k_{(\alpha)\ d}(x,x)$ 6 $\mathrm{vol}(\mathbb{B}^d(R))$1 2 $(2R)^{2\alpha}$，从而 $\overline{c}^2 \leqslant \mathrm{vol}(\mathbb{B}^d(R))2^{2\alpha-1}$。

注意，我们必须有 $\underline{c}\overline{c} \geqslant 1$。现在来证明这种等价性。

命题4（ $d \geqslant 1$的再生核希尔伯特空间范数）对于 $f \in L_2(\mathbb{B}^d(R))$，我们有:

$$\frac{1}{\underline{c}\overline{c}\sqrt{2}}\Omega_d^{(\alpha)(\mathrm{eq})}(f) \leqslant \Omega_d^{(\alpha)}(f) \leqslant 2\sqrt{2}\,\underline{c}\overline{c}\,\Omega_d^{(\alpha)(\mathrm{eq})}(f).$$

证明 对于上界，我们考虑一个函数 g 该函数达到定义 $\Omega_d^{(\alpha)(\mathrm{eq})}(f)^2$ 在公式（5）中给出的最小化问题。根据第5.1节，我们可以将 f 表示为 $f(x) = \int_{\mathbb{B}^d} (R) k_d^{(\alpha)}(x,y) d\lambda(y) + P(x)$，其中 λ 是定义在 $\mathbb{B}^d(R)$上的拉东测度，使得对于所有次数小于 α的多项式 Q，都有 $\int_{\mathbb{B}^d} (R) Q(y) d\lambda(y) = 0$。此外，由于我们有一个最小范数表示，因此可得，

$$\begin{aligned}\frac{1}{b_d^{(\alpha)}} \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \|\omega\|_2^{d+1+2\alpha} |\hat{g}(\omega)|^2 d\omega &\geqslant \int_{\mathbb{B}^d(R)} \int_{\mathbb{B}^d(R)} E_d^{(\alpha)}(x-y) d\lambda(y) d\lambda(x) \\ &= \int_{\mathbb{B}^d(R)} \int_{\mathbb{B}^d(R)} k_d^{(\alpha)}(x,y) d\lambda(y) d\lambda(x).\end{aligned}$$

由于核的再生性质，最后这个量等于 $\Omega_d^{(\alpha)}(f-P)^2$。

由于 $k^{(\alpha)}{}_{,d}{}^{(\mathrm{pol})}(x,y)$这一部分，多项式 P 可以在再生核希尔伯特空间中表示。因此， f 属于再生核希尔伯特空间。于是，我们只需证明 P 的再生核希尔伯特空间范数小于其在 $\mathbb{B}^d(R)$上的 L_2-范数，因为这样 f 的再生核希尔伯特空间范数就小于 $\Omega_d^{(\alpha)(\mathrm{eq})}(f)$的某个常数倍。

P 在 $\mathbb{B}^d(R)$上的 L_2-范数小于 g 的 L_2-范数（也就是 f的范数）与函数 $x \mapsto \int_{\mathbb{B}^d} (R) k_d^{(\alpha)}(x,y) d\lambda(y)$的 L_2-范数之和，而后者根据 $\overline{c}$的定义小于$\overline{c} R^\alpha \Omega_d^{(\alpha)}(f-P)$。

因此，根据 $\underline{c}$ 的定义，$\Omega_d^{(\alpha)}(P) \leqslant \frac{\underline{c}}{R^\alpha} \|P\|_{L_2(\mathbb{B}^d(R))}$) </style}，我们得到：

$$\begin{aligned}\Omega_d^{(\alpha)}(f) &\leqslant \Omega_d^{(\alpha)}(P) + \Omega_d^{(\alpha)}(f-P) \leqslant \frac{\underline{c}}{R^\alpha} \|P\|_{L_2(\mathbb{B}^d(R))} + \Omega_d^{(\alpha)}(f-P) \\ &\leqslant \frac{\underline{c}}{R^\alpha} \Big(\|f\|_{L_2(\mathbb{B}^d(R))} + \overline{c} R^\alpha \Omega_d^{(\alpha)}(f-P) \Big) + \Omega_d^{(\alpha)}(f-P) \\ &\leqslant \frac{\underline{c}}{R^\alpha} \|f\|_{L_2(\mathbb{B}^d(R))} + \underline{c}\overline{c} \Omega_d^{(\alpha)}(f-P) + \Omega_d^{(\alpha)}(f-P) \\ &\leqslant 2\underline{c}\overline{c} \left(\frac{1}{b_d^{(\alpha)}} \frac{1}{(2\pi)^d R} \int_{\mathbb{R}^d} \|\omega\|_2^{d+1+2\alpha} |\hat{f}(\omega)|^2 d\omega \right)^{1/2} + \frac{\underline{c}}{R^\alpha} \|f\|_{L_2(\mathbb{B}^d(R))}.\end{aligned}$$

因此，$\Omega_d^{(\alpha)}(f)^2 \leqslant 8\underline{c}^2 \overline{c}^2 \Omega_d^{(\alpha)(\mathrm{eq})}(f)^2$。

对于下界，给定 RKHS 中的 $f = \int_{\mathbb{B}^d} (R) k_d^{(\alpha)}(x,y) d\lambda(y)$，使平方范数 g that minimizes the squared norm $\frac{1}{b_d^{(\alpha)}} \frac{1}{2\pi}()^d \int_{\mathbb{R}^d} \|\omega\|_2^{d+1+2\alpha} |g(\omega)|^2 d\omega$ 最小的延拓，是可以写成 $g = \int_{\mathbb{B}^d} (R) k_d^{(\alpha)}(x,y) d\mu(y)$ 的延拓，其中 μ 与多项式正交，并且 $\int_{\mathbb{B}^d} (R)) \int_{\mathbb{B}^d(R))} k_d^{(\alpha)}(x,y) d\mu(x) d\mu(y)$ 最小。通过引入将 λ 测度 λ 投影到与所有次数小于 α 的多项式正交空间上所得的¯测度，我们得到：

$$\begin{aligned}\int_{\mathbb{B}^d(R))} \int_{\mathbb{B}^d(R))} k_d^{(\alpha)}(x,y) d\mu(x) d\mu(y) &\leqslant \int_{\mathbb{B}^d(R))} \int_{\mathbb{B}^d(R))} k_d^{(\alpha)}(x,y) d\bar{\lambda}(x) d\bar{\lambda}(y) \\ &\leqslant \int_{\mathbb{B}^d(R))} \int_{\mathbb{B}^d(R))} k_d^{(\alpha)}(x,y) d\lambda(x) d\lambda(y),\end{aligned}$$

其等于 $\Omega_d^{(\alpha)}(f)^2$. 因此，我们得到：

$$\begin{aligned}\Omega_d^{(\alpha)(\mathrm{eq})}(f)^2 &\leqslant \frac{1}{b_d^{(\alpha)}}\frac{1}{(2\pi)^d R}\int_{\mathbb{R}^d}\|\omega\|_2^{d+1+2\alpha}|\hat{g}(\omega)|^2 d\omega + \frac{\underline{c}^2}{R^{2\alpha}}\|g\|^2_{L_2(\mathbb{B}^d(R))}\\ &\leqslant \Omega_d^{(\alpha)}(f)^2 + \overline{c}^2\underline{c}^2\Omega_d^{(\alpha)}(f)^2 \leqslant 2\overline{c}^2\underline{c}^2\Omega_d^{(\alpha)}(f)^2,\end{aligned}$$

这就得到了所需的范数等价关系。 ■

5.3 当 $\alpha = 0$ = 0 时两种相互竞争的随机特征展开

我们首先可以考虑命题 3 中由神经网络得到的随机特征展开，也可以考虑基于傅里叶变换 [5] 的经典展开。事实上，对于 w 在球面上均匀分布的 $\mathbb{S}^d$，有：

$$\begin{aligned}k_d^{(\alpha)}(x,y) &= \mathbb{E}_w\big[k_1^{(\alpha)}(w^\top x, w^\top y)\big]\\ &= k_d^{(\alpha),(\mathrm{pol})}(x,y) + \frac{1}{R}\mathbb{E}_w\big[c_1^{(\alpha)}|w^\top(x-y)|^{2\alpha+1}\big] = \frac{1}{2} - \frac{1}{4R}\mathbb{E}_w\big[|w^\top(x-y)|\big],\end{aligned}$$

对于 $\alpha = 0$。由于 $|w^\top(x-y)| \leqslant 2R$ 几乎处处成立（因为 $x, y \in [-R, R]$），对于 $\alpha = 0$，我们可以使用该函数的一维傅里叶变换：

$$\varphi : u \mapsto (\frac{1}{2} - \frac{1}{4R}|u|)1_{|u|\leqslant 2R},$$

它等于

$$\hat{\varphi}(\omega) = \frac{\sin^2(R\omega)}{R\omega^2}.$$

因此，对于 $\|x-y\|_2 \leqslant 2R$，有

$$\begin{aligned}k_d^{(0)}(x,y) &= \mathbb{E}_w\Big[\frac{1}{2\pi R}\int_{-\infty}^{+\infty}\frac{\sin^2(R\tau)}{R\tau^2}e^{i\tau w^\top(x-y)}d\tau\Big]\\ &= \frac{2}{\mathrm{vol}(\mathbb{S}^{d-1})}\frac{1}{2\pi}\int_{\mathbb{S}^{d-1}}\int_0^{+\infty}\frac{\sin^2(R\tau)}{R\tau^{d+1}}e^{i\tau w^\top(x-y)}\tau^{d-1}d\tau dw\\ &= \frac{2}{\mathrm{vol}(\mathbb{S}^{d-1})}\frac{1}{2\pi R}\int_{\mathbb{R}^d}\frac{\sin^2(R\|\omega\|_2)}{\|\omega\|_2^{d+1}}e^{i\omega^\top(x-y)}d\omega \text{ using the change of variable } \omega = \tau w,\\ &= \frac{\Gamma(d/2)}{\pi^{d/2}}\frac{1}{2\pi R}\int_{\mathbb{R}^d}\frac{\sin^2(R\|\omega\|_2)}{\|\omega\|_2^{d+1}}e^{i\omega^\top(x-y)}d\omega\\ &= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\frac{\Gamma(d/2)(2\pi)^{d-1}}{\pi^{d/2}R}\frac{\sin^2(R\|\omega\|_2)}{\|\omega\|_2^{d+1}}e^{i\omega^\top(x-y)}d\omega.\end{aligned}$$

换句话说，函数 $x \mapsto (\frac{1}{2} - \frac{1}{R}c_d^{(0)}\|x\|_2)1_{\|x\|_2\leqslant 2R}$ 的傅里叶变换等于 $\omega \mapsto \Gamma(d/2)(2\pi)^{d-1}正弦^2(R\|\omega\|_2)\Gamma\|\omega\|_2^{d+1}$。由此自然得到随机特征余弦$(\omega^\top x + b)$，其中 b 在 $[-\pi, \pi]$ 上均匀分布，而 ω 从密度为 $2(2\pi)^d(d/2)(2\pi)^{d-1})正弦^2(R\|\omega\|_2)\|\omega\|_2^{d+1}$ 的分布中采样，这对应于

到 $\omega = \tau w$，其中 w 在球面上均匀分布，而 τ 从正弦^2 $(R\tau)$ $\pi R\tau^2$ 中采样，这可以通过从柯西分布中采样并使用拒绝采样来实现。我们也可以考虑 b 取值为 0 且 $\pi/2$ 均匀分布的情形，也就是说，使用随机特征余弦$(\omega^\top x)$和正弦 $(\omega^\top x)$。

经验比较： $d = 1$。我们首先在图1中对两个随机特征展开进行直观比较，然后在图2中进行数值比较，结果表明，基于神经网络的随机特征展开具有更好的逼近性质。[1]

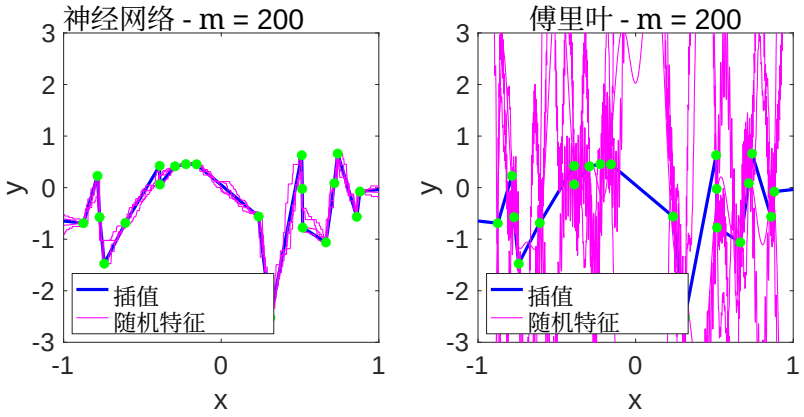


图1：利用完整再生核希尔伯特空间（蓝色）和随机特征展开，对绿色点进行最小范数插值，其中 $m = 200$。左：神经网络展开；右：傅里叶展开。图中绘出了随机特征的四次不同抽样结果。

$d = 1$ 的杠杆分数比较。我们想比较以下形式的两种随机特征展开：

$$k(x, y) = \mathbb{E}_v[\varphi(x, v)\varphi(y, v)],$$

对于特征 $\varphi(x, v) \in \mathbb{R}$ 对于 $v \in \mathcal{V}$。如 [12, 第 4], 节所述，为了评估随机特征展开逼近初始函数空间的能力，一个关键量是“杠杆分数”

$$v \mapsto \langle \varphi(\cdot, v), (\Sigma + \lambda I)^{-1}\varphi(\cdot, v)\rangle_{L_2(\mathbb{B}^d(R))},$$

其中 $\Sigma = \mathbb{E}_v\left[\varphi(\cdot, v) \otimes_{L_2(\mathbb{B}^d(R))} \varphi(\cdot, v)\right]$ 是作用于 $L_2(\mathbb{B}^d(R))$ 上的积分算子。 $v \in \mathcal{V}$ 上的最大杠杆分数直接影响所需随机特征的数量，以便在 L_2-范数下对原再生核希尔伯特空间的再生核希尔伯特空间球实现 λ-逼近：根据 [12, 命题 1], ，除对数项外，最大杠杆分数与所需随机特征的数量 m 成正比。

在附录C中，我们明确计算了这些 $d = 1$的杠杆分数，并发现对于神经网络特征，最大杠杆分数会在 $1/\sqrt{\lambda}$ 趋于零时 λ 发散；而对于随机傅里叶特征，它们发散得更快，达到 $1/\lambda$，这解释了上述实验结果为何显示出其优越性。

[1]可用于复现所有图的 Matlab 代码可在此处获取：https://www.di.ens.fr/~fbach/neural_样条_online.zip。

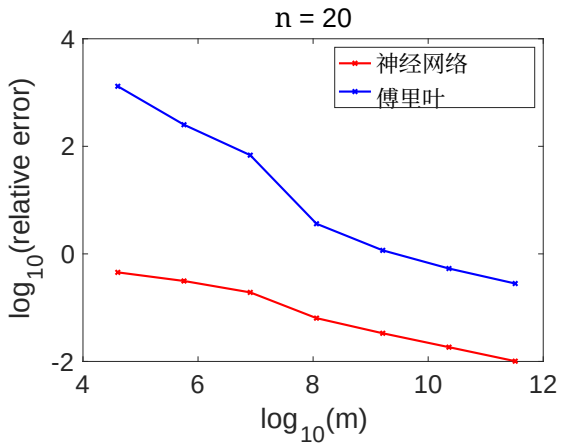


图2：对于不同的 m（随机特征数量），通过随机特征展开估计完整再生核希尔伯特空间（RKHS）的最小插值。我们从高斯分布中均匀抽取 $[-1,1]$ 中的 $n=20$ 个点 x_1、……、x_n，以及 n 个随机标签 y_1、……、y_n。然后，我们比较最小插值拟合与完整核及随机特征近似的 L_2（$[-1,1]$）范数。对于输入点的选择，我们对 20 次重复试验取平均；对于标签，则进行无限次重复试验（因为期望可以通过闭式形式求得）：给定测试点 x'_1、……、x'_m，以及训练和测试核矩阵 K 和 K'，连同它们的近似值 $\hat{K}$ 和 $\hat{K}'$，误差与 $\|K'K^{-1}y-\hat{K}'\hat{K}^{-1}y\|_2^2$ 成正比，因此我们可以计算关于 y 的期望，其值等于 $\|K'K^{-1}-\hat{K}'\hat{K}^{-1}\|_F^2$，这就是上图所绘制的量。

5.4 适用于所有 α 的新随机特征展开

对于所有 α，我们提出了一种新的核 $k_d^{(\alpha)}$，使经典多元样条成为正定核，并给出了随机特征展开，可用于高效估计。分析请参见 [7]。

其他核会导出与同一 Sobolev 范数等价的再生核希尔伯特空间范数，例如 Mat´ern 核 [23, 第 84] 页。这些核具有自然的随机傅里叶特征展开（其经验表现与 $\alpha=0$ 上文所示的相同），而我们的核基于神经网络，在用于随机特征展开时表现更佳。

6 结论

本文为与样条相关的核提供了新的随机特征展开。与现有基于傅里叶变换的展开相比，这些展开在欧几里得球上的索伯列夫空间具有更好的性质。正如 [14] 使用基于球谐函数的特征展开所做的那样，这种联系可用于为含有大量神经元的神经网络提供显式的逼近误差界（其中输入权重也经过估计）。

致谢

我们感谢 Nicolas Le Roux 和 Alessandro Rudi 就这项工作进行了富有启发性的讨论。我们感谢法国政府通过法国国家研究署管理的“未来投资计划”提供支持，项目编号为 ANR-19-P3IA-0001（PRAIRIE 3IA研究所）；同时也感谢欧洲研究委员会的支持（资助项目 SEQUOIA 724063）。

A 球面上均匀分布的若干引理

如果 w 在单位球面上服从均匀分布，则：

$$\begin{aligned}
\mathbb{E}[|w^\top z|^{2\alpha+1}] &= \|z\|_2^{2\alpha+1}\frac{\Gamma(1+\alpha)\Gamma(\frac{d}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{d}{2}+\frac{1}{2}+\alpha)} \\
\mathbb{E}[(w^\top z)^{2\alpha}] &= \|z\|_2^{2\alpha}\frac{\Gamma(\frac{1}{2}+\alpha)\Gamma(\frac{d}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{d}{2}+\alpha)} \\
\mathbb{E}[(w^\top z)^2] &= \|z\|_2^2/d \\
\mathbb{E}[z^\top w w^\top t] &= \frac{1}{d}z^\top t \\
\mathbb{E}[(z^\top w w^\top t)^2] &= \frac{1}{d(d+2)}\big[2(z^\top t)^2 + z^\top z \cdot t^\top t\big].
\end{aligned}$$

这可由 w_1^2服从参数为 $(\frac{1}{2}, \frac{d-1}{2})$ 的 Beta分布，并利用旋转不变性得到。 –

B 再生核希尔伯特空间范数表达式的证明， $d = 1$

B.1 命题的证明1

我们有 $\int_{-R}^{R}(x-b)_+^\alpha(y-b)_+^\alpha db = \int_{R}^{\text{最小值}\{x,y\}}(x-b)^\alpha(y-b)^\alpha db$，可以用 $s = \frac{x+y}{2}$ 和 $\delta = \frac{x-y}{2}$ 重新表述，从而得到 $x = s+\delta$ 和 $y = s-\delta$，其中 $\min\{x,y\} = s - |\delta|$。于是得到：

$$\begin{aligned}
\int_{-R}^{R}(x-b)_+^\alpha(y-b)_+^\alpha db &= \int_{-R}^{s-|\delta|}(s+\delta-b)^\alpha(s-\delta-b)^\alpha db = \int_{-R}^{s-|\delta|}((s-b)^2-\delta^2)^\alpha db \\
&= \int_{-R}^{s-|\delta|}\sum_{i=0}^{\alpha}\binom{\alpha}{i}(s-b)^{2i}(-1)^{i-\alpha}\delta^{2\alpha-2i}db \\
&= \sum_{i=0}^{\alpha}\binom{\alpha}{i}\frac{1}{2i+1}\Big[(R+s)^{2i+1}-|\delta|^{2i+1}\Big](-1)^{i-\alpha}\delta^{2\alpha-2i} \\
&= \sum_{i=0}^{\alpha}\binom{\alpha}{i}\frac{(-1)^{i-\alpha}}{2i+1}(R+s)^{2i+1}\delta^{2\alpha-2i} - \sum_{i=0}^{\alpha}\binom{\alpha}{i}\frac{(-1)^{i-\alpha}}{2i+1}\times|\delta|^{2\alpha+1} \\
&= A_\alpha(x,y) - B_\alpha|x-y|^{2\alpha+1},
\end{aligned}$$

with

$$\begin{aligned} A_\alpha &= \sum_{i=0}^{\alpha}\binom{\alpha}{i}\frac{(-1)^{i-\alpha}}{2i+1}(R+s)^{2i+1}\delta^{2\alpha-2i} = \int_{-y}^{R}(x+b)^\alpha(y+b)^\alpha db \\ B_\alpha &= \frac{1}{2^{2\alpha+1}}\sum_{i=0}^{\alpha}\binom{\alpha}{i}\frac{(-1)^{i-\alpha}}{2i+1} = \frac{1}{2^{2\alpha+1}}\int_0^1\sum_{i=0}^{\alpha}\binom{\alpha}{i}(-1)^{i-\alpha}x^{2i}dx \\ &= \frac{(-1)^\alpha}{2^{2\alpha+1}}\int_0^1(1-x^2)^\alpha dx = \frac{(-1)^\alpha}{2^{2\alpha+2}}2^{2\alpha+1}\int_0^1 u^\alpha(1-u)^\alpha du = \frac{(-1)^\alpha}{2}\frac{\Gamma(\alpha+1)^2}{\Gamma(2\alpha+2)}, \end{aligned}$$

使用变量代换 $\frac{1+x}{2}=u$, $\frac{1-x}{2}=1-u$。利用对称性可得:

$$\begin{aligned} k_d^{(\alpha)}(x,y) &= \frac{1}{4R}\int_{-y}^{R}(x+b)^\alpha(y+b)^\alpha db + \frac{1}{4R}\int_{y}^{R}(-x+b)^\alpha(-y+b)^\alpha db \\ &\qquad -\frac{(-1)^\alpha}{4R}\frac{\Gamma(\alpha+1)^2}{\Gamma(2\alpha+2)}|x-y|^{2\alpha+1} \\ &= \frac{1}{4R}\int_{-y}^{R}(x+b)^\alpha(y+b)^\alpha db + \frac{1}{4R}\int_{-R}^{-y}(-x-b)^\alpha(-y-b)^\alpha db \\ &\qquad -\frac{(-1)^\alpha}{4R}\frac{\Gamma(\alpha+1)^2}{\Gamma(2\alpha+2)}|x-y|^{2\alpha+1} \\ &= \frac{1}{4R}\int_{-R}^{R}(x+b)^\alpha(y+b)^\alpha db - \frac{(-1)^\alpha}{4R}\frac{\Gamma(\alpha+1)^2}{\Gamma(2\alpha+2)}|x-y|^{2\alpha+1}. \end{aligned}$$

然后我们可以使用二项式公式进行展开。

B.2 命题2的证明

如果我们有如下表示，其中 f 具有 $\alpha+1$ 个连续导数:

$$\forall x\in[-R,R], f(x)=\frac{1}{4R}\int_{-R}^{R}\left[\eta_+(b)(x-b)_+^\alpha+\eta_-(b)(b-x)_+^\alpha\right]db,$$

那么通过求 $(\alpha+1)$ 阶导数，我们必然有:

$$f^{(\alpha+1)}(x)=\frac{\alpha!}{4R}\eta_+(x)+\frac{\alpha!}{4R}(-1)^{\alpha+1}\eta_-(x).$$

因此我们有:

$$\begin{aligned} \eta_+(x) &= \frac{2R}{\alpha!}f^{(\alpha+1)}(x)+c(x) \\ \eta_-(x) &= (-1)^{\alpha+1}\frac{2R}{\alpha!}f^{(\alpha+1)}(x)+(-1)^\alpha c(x) \end{aligned}$$

对于某个 f 函数 c： $[-R,R]\to\mathbb{R}$。根据带积分余项的泰勒公式，我们有 :

$$\begin{aligned}
f(x) &= \sum_{i=0}^{\alpha}\frac{f^{(i)}(-R)}{i!}(x+R)^i+\int_{-R}^{R}\frac{f^{(\alpha+1)}(b)}{\alpha!}(x-b)_+^{\alpha}db\\
f(x) &= \sum_{i=0}^{\alpha}\frac{(-1)^i f^{(i)}(R)}{i!}(R-x)^i-(-1)^{\alpha}\int_{-R}^{R}\frac{f^{(\alpha+1)}(b)}{\alpha!}(b-x)_+^{\alpha}db,\ \text{and by averaging them,}\\
f(x) &= \frac{1}{2}\sum_{i=0}^{\alpha}\Big[\frac{f^{(i)}(-R)}{i!}(x+R)^i+\frac{(-1)^i f^{(i)}(R)}{i!}(R-x)^i\Big]\\
&\qquad +\frac{1}{2}\int_{-R}^{R}\frac{f^{(\alpha+1)}(b)}{\alpha!}\big[(x-b)_+^{\alpha}-(-1)^{\alpha}(b-x)_+^{\alpha}\big]db.
\end{aligned}$$

根据我们对 η_+ 和 η_- 的表达式，这意味着对所有 $x\in[-R,R]$

$$\begin{aligned}
\frac{1}{2}\sum_{i=0}^{\alpha}\Big[\frac{f^{(i)}(-R)}{i!}(x+R)^i+\frac{(-1)^i f^{(i)}(R)}{i!}(R-x)^i\Big] &= \frac{1}{4R}\int_{-R}^{R}\big[c(b)(x-b)_+^{\alpha}+(-1)^{\alpha}c(b)(b-x)_+^{\alpha}\big]db\\
&= \frac{1}{4R}\int_{-R}^{R}c(b)(x-b)^{\alpha}db,
\end{aligned}$$

从而对 $\int_{-R}^{R}c$（b）b^i 在 $i\in\{0,\dots,\alpha\}$ 的条件施加约束。通过最小化以下内容，可以得到最优的 c：

$$\frac{1}{4R}\int_{-R}^{R}\big[\eta_+(b)^2+\eta_-(b)^2\big]db=\frac{2R}{\alpha!^2}\int_{-R}^{R}f^{(\alpha+1)}(b)^2db+\frac{1}{2R}\int_{-R}^{R}c(b)^2db.$$

因此， c 必须是一个次数小于 α 的多项式，其系数是 $i\in\{0,\dots,\alpha\}$ 对应的 $f^{(i)}$（$\pm R$）的线性组合。这就得到了所需的结果。

B.3 一维多项式核

G给定一个定义在 $\mathbb{R}$ 上、次数小于 α（或等于）的多项式 P，如果我们可以将其写成 S:

$$P(x)=\frac{1}{2R}\int_{(\overline{\text{pol}})_1^R}\eta(b)(x-b)^{\alpha}db, \tag{6}$$

则其平方再生核希尔伯特空间范数（当 $k^{(\alpha)}$ 时）等于 $\frac{1}{4R}\int_{-R}^{R}\eta$（$b$）$^2db$ 的下确界。根据式（6）中的表示，我们有：

$$P^{(k)}(0)=\frac{1}{2R}\frac{\alpha!}{(\alpha-k)!}\int_{-R}^{R}\eta(b)(-b)^{\alpha-k}db=\frac{(-1)^{\alpha-k}}{2R}\frac{\alpha!}{(\alpha-k)!}\int_{-R}^{R}\eta(b)b^{\alpha-k}db,$$

它等于 $(-1)^{\alpha-k}\ \underline{\alpha!}\ \underline{(\alpha-k)!}\ \langle\eta,Q_{\alpha-k}\rangle_{L_2(\mathbb{B}^d(R))}$，其中 $Q_j(b)=b^j$。因此，既然我们要最小化 $\langle\eta,\eta\rangle_{L_2(\mathbb{B}^d(R))}$，解必定是一个多项式 $\eta=\sum_{j=0}^{\alpha}s_jQ_j$，其中 $s\in\mathbb{R}^{\alpha+1}$ 使其最小

$$\sum_{i,j=0}^{\alpha}s_is_j\langle Q_i,Q_j\rangle_{L_2(\mathbb{B}^d(R))}$$

满足 $(-1)^j 2R^j \alpha! \underline{P_{(\alpha-j)}}(0) = \sum_{i=0}^{\alpha} s_i \langle Q_i, Q_j \rangle_{L_2(\mathbb{B}^d(R))}$ 对所有 $j \in \{0, \ldots, \alpha\}$ 都成立。若 $P = \sum_{j=0}^{\alpha} t_j Q_j$，则得到 $(-1)^j 2R \binom{\alpha}{j}^{-1} t_{\alpha-j} = \sum_{i=0}^{\alpha} s_i \langle Q_i, Q_j \rangle_{L_2(\mathbb{B}^d(R))}$。由于单项式的格拉姆矩阵可逆，最优的 s 是系数 t 的线性函数。因此， P 的范数是系数的正定二次型。由此，该范数等价于次数小于 α （或等于）的多项式空间上的 L_2-范数。

B.4 维度为 $d \geqslant 1$ 时的多项式核

我们可以采用与上一节相同的推理，并需要证明我们能够将所有次数小于 α 的多项式表示为

$$P(x) = \frac{1}{2R} \int_{-R}^{R} \int_{\mathbb{S}^d} \eta(w, b)(w^\top x + b)^\alpha dw db,$$

对于 $\eta(w,b)$，平方可积。 对 $x = 0$ 处的所有偏导数取值后，这就要求 a ll

$$\int_{-R}^{R} \int_{\mathbb{S}^d} \eta(w, b) w_1^{u_1} \cdots w_d^{u_d} b^v dw db$$

对于 $u_1 + \cdots + u_d + v = \alpha$，这些值是固定的；并且由于多项式族 $(w, b) \mapsto w_1^{u_1} \cdots w_d^{u_d} b^v$ 在线性空间 $L_2(\mathbb{S}^d \times [-1, 1])$ 中线性无关，沿用上述推理可得一个再生核希尔伯特空间范数，它等价于次数小于 α （或等于）的多项式空间上的 L_2-范数。

C 计算杠杆分数

在本节中，我们将明确计算这两种展开中 $d = 1$ 和 $\alpha = 0$ 的杠杆分数。

C.1 一般解

我们考虑 $d = 1, \quad R = 1$，以及关于 $[-1, 1$上的均匀分布的经典积分算子 $]$:

$$\Sigma f(x) = \frac{1}{2} \int_{-1}^{1} k_1^{(0)}(x, y) f(y) dy = \frac{1}{4} \int_{-1}^{1} f(y) dy - \frac{1}{8} \int_{-1}^{1} |x - y| f(y) dy.$$

给定函数 $g \in L_2([-1, 1])$，我们的目标是计算杠杆分数:

$$\frac{1}{2} \int_{-1}^{1} g(x) \left[(\Sigma + \lambda I)^{-1} g \right](x) dx.$$

因此，我们计算 $f = (\Sigma + \lambda I)^{-1} g$，它满足:

$$g(x) = \frac{1}{4} \int_{-1}^{1} f(y) dy - \frac{1}{8} \int_{-1}^{1} |x - y| f(x) dy + \lambda f(x).$$

求两次导数　　并利用 $x \mapsto |x-y|$ 的二阶导数为 $2\delta_y$ 这一事实，我们得到　　t:

$$g''(x) = \lambda f''(x) - \frac{1}{4} f(x).$$

一旦我们知道上述常微分方程的一个解 f_0，就可以得到所有解：

$$f(x) = f_0(x) + A\cosh\frac{x}{2\sqrt{\lambda}} + B\sinh\frac{x}{2\sqrt{\lambda}}, \tag{7}$$

其中 A、$B \in \mathbb{R}$ 为任意常数。

求得一个“闭式”解。我们可以使用 f 标准方法求解 [24]，写成 $f(x) = e^{\frac{x}{2\sqrt{\lambda}}} a(x)$，于是

$$\begin{aligned} g''(x) &= \lambda e^{\frac{x}{2\sqrt{\lambda}}} a''(x) + \sqrt{\lambda} e^{\frac{x}{2\sqrt{\lambda}}} a'(x) \\ e^{-\frac{x}{2\sqrt{\lambda}}} g''(x) &= \lambda a''(x) + \sqrt{\lambda} a'(x). \end{aligned}$$

于是我们写作 $a'(x) = e^{-\frac{x}{\sqrt{\lambda}}} c(x)$，从而

$$e^{-\frac{x}{2\sqrt{\lambda}}} g''(x) = \lambda e^{-\frac{x}{\sqrt{\lambda}}} c'(x),$$

因此

$$c'(x) = \frac{1}{\lambda} e^{\frac{x}{2\sqrt{\lambda}}} g''(x),$$

从而通过积分得到一个特解：

$$c(x) = \frac{1}{2\lambda} \int_{-1}^{1} e^{\frac{y}{2\sqrt{\lambda}}} g''(y) \operatorname{sign}(x-y) dy.$$

此外，我们得到一个特解：

$$\begin{aligned} a(x) &= \frac{1}{2} \int_{-1}^{1} e^{-\frac{y}{\sqrt{\lambda}}} c(y) \operatorname{sign}(x-y) dy \\ &= \frac{1}{4\lambda} \int_{-1}^{1} e^{-\frac{y}{\sqrt{\lambda}}} \int_{-1}^{1} e^{\frac{t}{2\sqrt{\lambda}}} g''(t) \operatorname{sign}(y-t) dt \operatorname{sign}(x-y) dy \\ f(x) &= \frac{1}{4\lambda} \int_{-1}^{1} \int_{-1}^{1} e^{\frac{x}{2\sqrt{\lambda}}} e^{-\frac{y}{\sqrt{\lambda}}} e^{\frac{t}{2\sqrt{\lambda}}} g''(t) \operatorname{sign}(y-t) \operatorname{sign}(x-y) dy dt, \end{aligned}$$

所有解都可以通过加上 $A\cosh\frac{x}{2\sqrt{\lambda}} + B\sinh\frac{x}{2\sqrt{\lambda}}$ 得到。

求　ng 常数 A 和 B。对于 $g = (\Sigma + \lambda I) f$ 的唯一解 f，我们有　　：

$$\begin{aligned} g(1) &= \frac{1}{4}\int_{-1}^{1} f(y)dy - \frac{1}{8}\int_{-1}^{1}(1-y)f(y)dy + \lambda f(1) \\ g(-1) &= \frac{1}{4}\int_{-1}^{1} f(y)dy - \frac{1}{8}\int_{-1}^{1}(1+y)f(y)dy + \lambda f(-1), \text{ leading to} \\ g(1)+g(-1) &= \frac{1}{4}\int_{-1}^{1} f(y)dy + \lambda[f(1)+f(-1)] \qquad (8) \\ g(1)-g(-1) &= \frac{1}{4}\int_{-1}^{1} y f(y)dy + \lambda[f(1)-f(-1)]. \qquad (9) \end{aligned}$$

For $f(x)=\cosh\frac{x}{2\sqrt{\lambda}}$, we have: $\int_{-1}^{1} f(x)dx = 4\sqrt{\lambda}\sinh\frac{1}{2\sqrt{\lambda}}$.

For $f(x)=\sinh\frac{x}{2\sqrt{\lambda}}$, we have: $\int_{-1}^{1} xf(x)dx = 4\sqrt{\lambda}\cosh\frac{1}{2\sqrt{\lambda}} - 8\lambda\sinh\frac{1}{2\sqrt{\lambda}}$.

因此，对于式（7）中的解：

$$\begin{aligned} g(1)+g(-1) &= \frac{1}{4}\int_{-1}^{1} f_0(y)dy + \lambda[f_0(1)+f_0(-1)] + A\Big[\sqrt{\lambda}\sinh\frac{1}{2\sqrt{\lambda}} + 2\lambda\cosh\frac{1}{2\sqrt{\lambda}}\Big] \\ g(1)-g(-1) &= \frac{1}{4}\int_{-1}^{1} yf_0(y)dy + \lambda[f_0(1)-f_0(-1)] + B\sqrt{\lambda}\cosh\frac{1}{2\sqrt{\lambda}}. \end{aligned}$$

因此，要得到 A 和 B，我们只需计算 $f_0(1)$、$f_0(-1)$以及 $\int_{-1}^{1} f_0(y)dy$ 和

$\int_{-1}^{1} yf_0(y)dy$.

C.2 神经网络

我们考虑 $g(x)=1_{x>b}=(x-b)_+^0$，然后考虑 $f_0(x)$ 余弦双曲函数$\frac{x-b}{2\sqrt{\lambda}}$。我们有：

$$\begin{aligned} g(x)-\lambda f_0(x) &= 1_{x>b}\Big[1-\cosh\frac{x-b}{2\sqrt{\lambda}}\Big] \\ g'(x)-\lambda f_0'(x) &= -\frac{1}{2\sqrt{\lambda}}1_{x>b}\sinh\frac{x-b}{2\sqrt{\lambda}} \\ g''(x)-\lambda f_0''(x) &= -\frac{1}{4\lambda}1_{x>b}\cosh\frac{x-b}{2\sqrt{\lambda}} = -\frac{1}{4}f_0(x), \end{aligned}$$

因此 f_0 是一个特解。我们有：

$$\begin{aligned} f_0(1)+f_0(-1) &= f_0(1)-f_0(-1)=f_0(1)=\frac{1}{\lambda}\cosh\frac{1-b}{2\sqrt{\lambda}} \\ \int_{-1}^{1} f_0(x)dx &= \int_b^1 \frac{1}{\lambda}\cosh\frac{x-b}{2\sqrt{\lambda}}dx = \frac{2}{\sqrt{\lambda}}\sinh\frac{1-b}{2\sqrt{\lambda}} \\ \int_{-1}^{1} xf_0(x)dx &= \int_b^1 \frac{1}{\lambda}x\cosh\frac{x-b}{2\sqrt{\lambda}}dx = \int_b^1 \frac{1}{\lambda}b\cosh\frac{x-b}{2\sqrt{\lambda}}dx + \int_b^1 \frac{1}{\lambda}(x-b)\cosh\frac{x-b}{2\sqrt{\lambda}}dx \\ &= \frac{2b}{\sqrt{\lambda}}\sinh\frac{1-b}{2\sqrt{\lambda}} + \frac{2}{\sqrt{\lambda}}\Big[(1-b)\sinh\frac{1-b}{2\sqrt{\lambda}} - 2\sqrt{\lambda}\cosh\frac{1-b}{2\sqrt{\lambda}}\Big] \\ &= \frac{2}{\sqrt{\lambda}}\Big[\sinh\frac{1-b}{2\sqrt{\lambda}} - 2\sqrt{\lambda}\cosh\frac{1-b}{2\sqrt{\lambda}}\Big]. \end{aligned}$$

Thus

$$\begin{aligned} \frac{1}{2} &= \frac{1}{2\sqrt{\lambda}}\sinh\frac{1-b}{2\sqrt{\lambda}} + \cosh\frac{1-b}{2\sqrt{\lambda}} + 2\lambda A\Big[\frac{1}{2\sqrt{\lambda}}\sinh\frac{1}{2\sqrt{\lambda}} + \cosh\frac{1}{2\sqrt{\lambda}}\Big] \\ \frac{1}{2} &= \frac{1}{2\sqrt{\lambda}}\sinh\frac{1-b}{2\sqrt{\lambda}} + B\sqrt{\lambda}\cosh\frac{1}{2\sqrt{\lambda}}, \end{aligned}$$

由此可以解出 A和 B。此外

$$\begin{aligned}\int_{-1}^{1} f(x)g(x)dx &= \int_{-1}^{1} g(x)\big[f_0(x) + A\cosh\frac{x}{2\sqrt{\lambda}} + B\sinh\frac{x}{2\sqrt{\lambda}}\big]dx \\ &= \frac{2}{\sqrt{\lambda}}\sinh\frac{1-b}{2\sqrt{\lambda}} + 2A\sqrt{\lambda}\big[\sinh\frac{1}{2\sqrt{\lambda}} - \sinh\frac{b}{2\sqrt{\lambda}}\big] + 2B\sqrt{\lambda}\big[\cosh\frac{1}{2\sqrt{\lambda}} - \cosh\frac{b}{2\sqrt{\lambda}}\big] \\ &= \frac{2}{\sqrt{\lambda}}\sinh\frac{1-b}{2\sqrt{\lambda}} + \frac{1}{\sqrt{\lambda}}\frac{\big[\frac{1}{2} - \frac{1}{2\sqrt{\lambda}}\sinh\frac{1-b}{2\sqrt{\lambda}} - \cosh\frac{1-b}{2\sqrt{\lambda}}\big]}{\frac{1}{2\sqrt{\lambda}}\sinh\frac{1}{2\sqrt{\lambda}} + \cosh\frac{1}{2\sqrt{\lambda}}}\big[\sinh\frac{1}{2\sqrt{\lambda}} - \sinh\frac{b}{2\sqrt{\lambda}}\big] \\ &\quad + \frac{1 - \frac{1}{\sqrt{\lambda}}\sinh\frac{1-b}{2\sqrt{\lambda}}}{\cosh\frac{1}{2\sqrt{\lambda}}}\big[\cosh\frac{1}{2\sqrt{\lambda}} - \cosh\frac{b}{2\sqrt{\lambda}}\big],\end{aligned}$$

这就是我们所需的量（乘以 2）。该量在 $b = 0$ 处达到最大值，此时其值为：

$$\begin{aligned}&\frac{2}{\sqrt{\lambda}}\sinh\frac{1}{2\sqrt{\lambda}} + \frac{1}{\sqrt{\lambda}}\frac{\big[\frac{1}{2} - \frac{1}{2\sqrt{\lambda}}\sinh\frac{1}{2\sqrt{\lambda}} - \cosh\frac{1}{2\sqrt{\lambda}}\big]}{\frac{1}{2\sqrt{\lambda}}\sinh\frac{1}{2\sqrt{\lambda}} + \cosh\frac{1}{2\sqrt{\lambda}}}\big[\sinh\frac{1}{2\sqrt{\lambda}}\big] + \frac{1 - \frac{1}{\sqrt{\lambda}}\sinh\frac{1}{2\sqrt{\lambda}}}{\cosh\frac{1}{2\sqrt{\lambda}}}\big[\cosh\frac{1}{2\sqrt{\lambda}} - 1\big] \\ = \;& \frac{1}{\sqrt{\lambda}}\frac{\big[\frac{1}{2} + \frac{1}{2\sqrt{\lambda}}\sinh\frac{1}{2\sqrt{\lambda}} - \cosh\frac{1}{2\sqrt{\lambda}}\big]}{\frac{1}{2\sqrt{\lambda}}\sinh\frac{1}{2\sqrt{\lambda}} + \cosh\frac{1}{2\sqrt{\lambda}}}\big[\sinh\frac{1}{2\sqrt{\lambda}}\big] + \big[1 - \frac{1}{\sqrt{\lambda}}\sinh\frac{1}{2\sqrt{\lambda}}\big]\cdot\big[1 - \frac{1}{\cosh\frac{1}{2\sqrt{\lambda}}}\big] \\ = \;& \Big[1 + \frac{1/2}{\frac{1}{2\sqrt{\lambda}}\sinh\frac{1}{2\sqrt{\lambda}} + \cosh\frac{1}{2\sqrt{\lambda}}}\Big]\frac{1}{\sqrt{\lambda}}\sinh\frac{1}{2\sqrt{\lambda}} + \big[1 - \frac{1}{\sqrt{\lambda}}\sinh\frac{1}{2\sqrt{\lambda}}\big]\cdot\big[1 - \frac{1}{\cosh\frac{1}{2\sqrt{\lambda}}}\big] \\ = \;& \frac{\frac{1}{2\sqrt{\lambda}}\sinh\frac{1}{2\sqrt{\lambda}}}{\frac{1}{2\sqrt{\lambda}}\sinh\frac{1}{2\sqrt{\lambda}} + \cosh\frac{1}{2\sqrt{\lambda}}} + \big[1 - \frac{1}{\cosh\frac{1}{2\sqrt{\lambda}}}\big] + \frac{1}{\sqrt{\lambda}}\frac{\sinh\frac{1}{2\sqrt{\lambda}}}{\cosh\frac{1}{2\sqrt{\lambda}}}.\end{aligned}$$

因此，最大杠杆分数的阶为 $\frac{1}{2\sqrt{\lambda}}$。

C.3 傅里叶特征

我们考虑 $g(x) = e^{i\omega x}$，以便同时得到余弦 ωx 和正弦 ωx。然后我们可以取 $f_0(x) = \frac{\omega^2}{\lambda\omega^2 + \frac{1}{4}}e^{i\omega x}$ 作为一个特殊解，因为

$$\lambda f_0''(x) - \frac{1}{4}f_0(x) = \omega^2\frac{-\lambda\omega^2 - \frac{1}{4}}{\lambda\omega^2 + \frac{1}{4}}e^{i\omega x} = g''(x).$$

由公式（8）和公式（9）可得：

$$\begin{aligned}2\cos\omega &= \frac{\omega^2}{4\lambda\omega^2 + 1}\frac{1}{i\omega}2i\sin\omega + \frac{\lambda\omega^2}{\lambda\omega^2 + \frac{1}{4}}2\cos\omega + A\Big[\sqrt{\lambda}\sinh\frac{1}{2\sqrt{\lambda}} + 2\lambda\cosh\frac{1}{2\sqrt{\lambda}}\Big] \\ 2i\sin\omega &= \frac{\omega^2}{4\lambda\omega^2 + 1}\Big[\frac{1}{\omega^2}e^{i\omega x}(1 - i\omega x)\Big]_{-1}^{1} + \frac{\lambda\omega^2}{\lambda\omega^2 + \frac{1}{4}}2i\sin\omega + B\sqrt{\lambda}\cosh\frac{1}{2\sqrt{\lambda}}.\end{aligned}$$

这得到常数 A 和 B 的显式公式:

$$\begin{aligned}\frac{2\cos\omega - 2\omega\sin\omega}{4\lambda\omega^2+1} &= A\Big[\sqrt{\lambda}\sinh\frac{1}{2\sqrt{\lambda}} + 2\lambda\cosh\frac{1}{2\sqrt{\lambda}}\Big]\\ \frac{2i\sin\omega}{4\lambda\omega^2+1} &= \frac{1}{4\lambda\omega^2+1}\Big[2i\sin\omega - 2i\omega\cos\omega\Big] + B\sqrt{\lambda}\cosh\frac{1}{2\sqrt{\lambda}},\ \text{leading to}\\ \frac{2i\omega\cos\omega}{4\lambda\omega^2+1} &= B\sqrt{\lambda}\cosh\frac{1}{2\sqrt{\lambda}}.\end{aligned}$$

于是我们得到

$$\begin{aligned}A &= \frac{2\cos\omega - 2\omega\sin\omega}{4\lambda\omega^2+1}\frac{1}{\sqrt{\lambda}\sinh\frac{1}{2\sqrt{\lambda}} + 2\lambda\cosh\frac{1}{2\sqrt{\lambda}}}\\ \frac{B}{i} &= \frac{2\omega\cos\omega}{4\lambda\omega^2+1}\frac{1}{\sqrt{\lambda}\cosh\frac{1}{2\sqrt{\lambda}}}.\end{aligned}$$

因此，$g(x)$ =余弦 ωx的解为 $f(x) = \frac{\omega^2}{\lambda\omega^2+\frac{1}{4}}$余弦 $\omega x + A$双曲余弦 $\frac{x}{2\sqrt{\lambda}}$，而$g(x)$ =正弦 ωx的解为 $f(x) = \frac{\omega^2}{\lambda\omega^2+\frac{1}{4}}$正弦 $\omega x + \frac{B}{i}$双曲正弦 $\frac{x}{2\sqrt{\lambda}}$。

因此，我们可以计算 $g(x)$ = 余弦ωx

$$\begin{aligned}\int_{-1}^{1} f(x)g(x)dx &= \int_{-1}^{1}\cos\omega x\Big[\frac{\omega^2}{\lambda\omega^2+\frac{1}{4}}\cos\omega x + A\cosh\frac{x}{2\sqrt{\lambda}}\Big]dx\\ &= \frac{\omega^2}{\lambda\omega^2+\frac{1}{4}}\Big(1+\frac{1}{2}\frac{\sin\omega}{\omega}\Big) + A\int_{-1}^{1}\cos\omega x\cosh\frac{x}{2\sqrt{\lambda}}dx\\ &= \frac{\omega^2}{\lambda\omega^2+\frac{1}{4}}\Big(1+\frac{1}{2}\frac{\sin\omega}{\omega}\Big) + \frac{2A}{\omega^2+\frac{1}{4\lambda}}\Big[\frac{1}{2\sqrt{\lambda}}\cos\omega\sinh\frac{1}{2\sqrt{\lambda}} + \omega\sin\omega\cosh\frac{1}{2\sqrt{\lambda}}\Big].\end{aligned}$$

对于 $g(x)$ = 正弦ωx，我们得到:

$$\begin{aligned}\int_{-1}^{1} f(x)g(x)dx &= \int_{-1}^{1}\sin\omega x\Big[\frac{\omega^2}{\lambda\omega^2+\frac{1}{4}}\sin\omega x + \frac{B}{i}\sinh\frac{x}{2\sqrt{\lambda}}\Big]dx\\ &= \frac{\omega^2}{\lambda\omega^2+\frac{1}{4}}\Big(1-\frac{1}{2}\frac{\sin\omega}{\omega}\Big) + \frac{B}{i}\int_{-1}^{1}\sin\omega x\sinh\frac{x}{2\sqrt{\lambda}}dx\\ &= \frac{\omega^2}{\lambda\omega^2+\frac{1}{4}}\Big(1-\frac{1}{2}\frac{\sin\omega}{\omega}\Big) + \frac{2B/i}{\omega^2+\frac{1}{4\lambda}}\Big[\frac{1}{2\sqrt{\lambda}}\sin\omega\cosh\frac{1}{2\sqrt{\lambda}} - \omega\cos\omega\sinh\frac{1}{2\sqrt{\lambda}}\Big].\end{aligned}$$

由此我们得到两个杠杆分数（除以 2）。我们注意到，当 ω 趋于无穷大时，这两个杠杆分数趋于1/(2λ)，这是所有 ω中的最大值。

C.4 经验比较

详见 [25, 附录 A], 我们可以从 $[-1,1]$ 中的网格估计杠杆分数，网格包含 n 个点，方法是计算$\sum_i^n{}_{,\ j=1}\varphi($ $x_i,\ \ v)\varphi(^{x_j,\ \ v)}\big[(K+n\lambda I)^{-1}\big]_{ij}$，并与上文给出的理论表达式进行比较，两者吻合。参见图3。

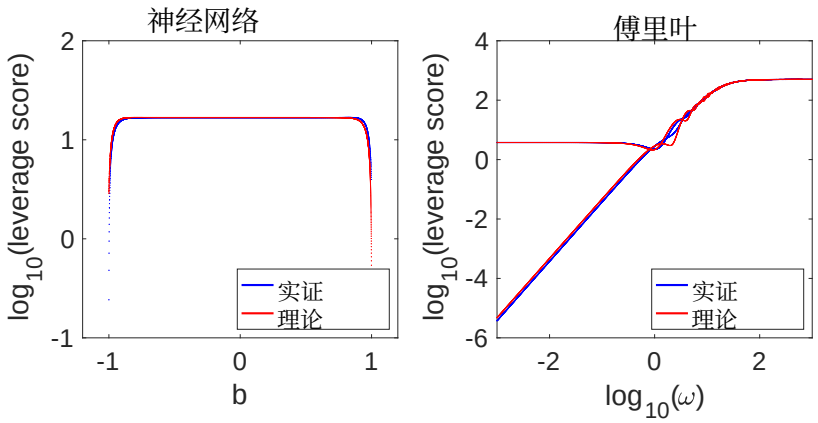


图 3：神经网络特征（左）和傅里叶特征（右）的经验杠杆分数与理论杠杆分数对比。我们使用了 $\lambda = 10^{-3}$ 和 $n = 4096$。

参考文献

[1] Trevor Hastie、Robert Tibshirani 和 Jerome H. Friedman。《统计学习要素：数据挖掘、推断与预测》。施普林格，2009年。（引自第 1 页）[2] Larry Wasserman。《非参数统计全书》。施普林格科学与商业媒体，2006年。（引自第 1 页）[3] L´aszl´o Gy¨orfi、Michael Kohler、Adam Krzyzak 和 Harro Walk。《无分布假设的非参数回归理论》。施普林格，2002年。（引自第 1 页）[4] Radford M. Neal。《神经网络的贝叶斯学习》。博士论文，多伦多大学，1995年。（引自第 1 页）[5] Ali Rahimi 和 Benjamin Recht。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年。（引自第 1、2 和 11 页）[6] Bernhard Sch¨olkopf 和 Alexander J. Smola。《核方法学习：支持向量机、正则化、优化及其他》。麻省理工学院出版社，2002年。（引自第 2 页）[7] Alessandro Rudi 和 Lorenzo Rosasco。基于随机特征学习的泛化性质。载于《神经信息处理系统进展》，第 3215–3225 页，2017年。（引自第 2 和 13 页）[8] Nicolas Le Roux 和 Yoshua Bengio。连续神经网络。载于人工智能与统计学，第 404–411 页，2007年。（引自第 2 和 3 页）[9] Jean Duchon。在索博列夫空间中最小化旋转不变半范数的样条。载于《多变量函数构造理论》，第 85–100 页。施普林格，1977年。（引自第 2 和 8 页）[10] Martin D. Buhmann。《径向基函数：理论与实现》，第 12 卷。剑桥大学出版社，2003 年。（引自第 2 和 8 页）[11] Vinod Nair 和 Geoffrey E. Hinton。修正线性单元提升受限玻尔兹曼机的性能。载于国际机器学习大会，2010年。（引自第 2 页）

[12] Francis Bach。关于核求积规则与随机特征展开之间的等价性。《机器学习研究杂志》，18(1):714–751，2017年。（引用见第2页和第12页）[13] Youngmin Cho 和 Lawrence K. Saul。用于深度学习的核方法。载于《神经信息处理系统进展》，2009年。（引用见第3页）[14] Francis Bach。利用凸神经网络打破维度灾难。《机器学习研究杂志》，18(1):629–681，2017年。（引用见第3、5和13页）[15] Meyer Scetbon 和 Zaid Harchaoui。点积核的谱分析。载于国际人工智能与统计学会议，第3394–3402页，2021年。（引用见第3页）[16] Amit Daniely、Roy Frostig 和 Yoram Singer。迈向对神经网络的深入理解：初始化的力量与关于表达能力的对偶视角。《神经信息处理系统进展》，29，2016年。（引用见第3页）[17] Agustinus Kristiadi、Matthias Hein 和 Philipp Hennig。一种无限特征延拓方法，可修正贝叶斯 ReLU 网络的渐近过度自信。《神经信息处理系统进展》，34，2021年。（引用见第4页）[18] Alain Berlinet 和 Christine Thomas-Agnan。《概率与统计中的再生核希尔伯特空间》。施普林格科学与商业媒体，2011年。（引用见第5页和第8页）[19] Robert A. Adams 和 John J. F. Fournier。《索博列夫空间》。学术出版社，2003年。（引用见第6页）[20] Holger Wendland。《散乱数据逼近》。剑桥大学出版社，2004年。（引用见第8页）[21] Grace Wahba。《观测数据的样条模型》。SIAM，1990年。（引用见第8页）[22] Friedrich Gerard Friedlander、Mark Suresh Joshi、M. Joshi 和 Mohan C. Joshi。《分布理论导论》。剑桥大学出版社，1998年。（引用见第8页）[23] Christopher K. I. Williams 和 Carl Edward Rasmussen。《用于机器学习的高斯过程》。麻省理工学院出版社，2006年。（引用见第2页和第13页）[24] Carl M. Bender、Steven Orszag 和 Steven A. Orszag。《科学家与工程师的高级数学方法 I：渐近方法与摄动理论》，第1卷。施普林格科学与商业媒体，1999年。（引用见第18页）[25] Edouard Pauwels、Francis Bach 和 Jean-Philippe Vert。利用正则化 Christoffel 函数关联杠杆分数与密度。《神经信息处理系统进展》，31，2018年。（引用见第21页）