arXiv:2301.01327v2 [cs.LG] 5 Jan 2023

算子理论、核与前馈神经网络

PALLE E.T. JORGENSEN、MYUNG-SIN SONG 和 JAM ES TIAN

谨以此文纪念 *Ka-Sing Lau* 教授。

摘要。本文展示了特定类别的正定核如何成为强有力的工具，用于分析多层前馈神经网络模型的迭代算法。我们重点研究一些特定的核，它们能够很好地适用于学习算法，以处理在尺度变换的前馈迭代中呈现内在自相似性的数据集／特征。

目录

1. 引言 12. 神经网络（NN）与再生核希尔伯特空间（RKHS） 22.1. 基底方法 92.2. 算子的对偶对 102.3. 函数与算子 122.4. 当.....时 $K = L$ 133. 基于正定核的神经网络激活函数 154. 分形图像的应用 18参考文献 21

1. 引言

近年来，许多作者为前馈神经网络（NN）算法提出了各种不同的方法 [ZC23,AK23,HL23,DWZ+23]，并提出了基于核的优化项。本文建立了一些算子理论方面的新结果，并将其应用于这一问题。前馈神经网络模型的应用涵盖多种机器学习场景，而基于核的深度神经网络 [GK23, MK23, BSW23, SHO22, GKNV22,GPR+21,Kut20,GKP20]。

前馈神经网络模型的一个共同特点是采用特定的预设迭代过程，其中包括 (i) ReLU 函数 [JR23、OSZ22、CKM22、CCK22、JR22、CL22], (ii) 从预设的仿射映射系统进行替换。此外，(iii) 每一步都会通过选择一个激活函数与下一步相连。本文表明，与前馈神经网络构造的三个步骤及其迭代相对应，存在自然的正定核。我们

2000 数学主题分类。41A30, 46E22, 47B32, 68T07, 92B20。关键词和短语。算法、乘子、谱分解、正规算子、迭代函数系统、分形测度、前馈神经网络、显式核、ReLU、再生核希尔伯特空间、正定核、复合算子。

认为，这由此为基于核的前馈神经网络模型提供了一种更直接的工具。我们方法的这一优势基于两个事实。首先，我们提出了一种直接的核迭代概念，可涵盖传统的函数论前馈神经网络步骤。其次，我们的方法提供了更直接、更自然的核选择，用以支配深度神经网络模型中的近似过程，例如图神经网络构造。

尽管正定核及其相关的再生核希尔伯特空间已广泛应用于纯数学与应用数学的各个领域，但本文将聚焦于核在前馈网络模型中的一种新作用。具体而言，本文的主要目的是介绍若干特定的正定核族，它们可作为分析多层前馈神经网络的强大工具。

一般而言，再生核构造及其相应的再生核希尔伯特空间是多种应用中的强大工具。在当前的核神经网络（KNNs）框架下，它们的作用可概括如下：从所要解决的问题出发，当我们通过迭代函数系迭代（例如，通过类似康托尔集的分形极限）构造再生核希尔伯特空间(μ) 时，类似康托尔集的 μ激活函数便成为再生核希尔伯特空间(μ) 的相对再生偶极函数，如下图 3.1 所示。

2. 神经网络（NN）与再生核希尔伯特空间（RKHS）

本文的一个主题是开发用于设计前馈神经网络构造的新工具。为此，我们指出了正定核及其相关生成函数在神经网络算法中的应用。这些基于核的函数包括更为熟悉的 ReLU 函数，参见下文定理3.3和3.4。我们强调，特定的再生核希尔伯特空间构造将在定理3.3 和3.4的意义下具有相对性，即内积再现函数值之差。

我们将核和函数用于前馈神经网络（NN）算法的方法，基于对两类算子的系统研究。它们的作用如下：（i）作用于预先给定的核希尔伯特空间之间；（ii）作用于网络中带索引的层级，即作用于固定层级内、因而处于所选核希尔伯特空间中的其他算子。情形（i）包括在核希尔伯特空间背景下对复合算子的系统研究（参见推论2.7 和 2.9）；情形（ii）则研究乘子算子及其伴随算子，例如参见定理 2.15。我们强调，下面讨论的两类运算取决于特定神经网络模型中各层所选用的核。总体而言，这些算子族能够在相关的广义多分辨率系统中实现黑箱滤波器项，包括由复合后乘法构成的算子。后续各节将介绍具体的三维应用，见第3 和4节。

约定。本文中，我们将研究函数希尔伯特空间，例如再生核希尔伯特空间（再生核希尔伯特空间）、 L^2 空间和 Sobolev空间。我们假设这些都是实值函数的希尔伯特空间。内积记作 $\langle\cdot,\cdot\rangle$，并用 $\langle\cdot,\cdot\rangle$ 的下标表示所考虑的希尔伯特空间。此外，我们所说的求导或微分算子，均指弱导数，即分布意义下的求导，或利用所考虑空间的自然对偶性。

我们在此将讨论范围限制于实值情形，是出于目前对前馈神经网络的应用需要。不过，下文第2 节中的许多一般性结果可推广到复再生核希尔伯特空间理论。后者对于研究复域的几何与势理论也十分重要，参见例如 [Eng96]。

核机器的优势部分源于以下事实。首先，核机器可将低维数据集（通常是非线性的）中的点映射到更高维空间。这个线性“超空间”的维数可以是无限的，但其设计旨在便于优化和高效进行特征编码。因此，核方法使我们能够针对手头的问题，通过再生核希尔伯特空间内积找到分离超平面的系数；每对高维特征对应选取一个内积。虽然多种类型的核机器已应用数十年，但正是支持向量机（SVMs）的发明使核如今成为关注焦点（例如参见 [CST01、PORSTS21、HSTHD11、HST10、RSSST06、STWCK05]）。如今，支持向量机已应用于各种领域，包括生物信息学（用于寻找不同蛋白质序列之间的相似性）、机器视觉和手写识别。深度神经网络（将在下文第 3 节和第 4 节讨论）由人工神经元层构成：输入层、输出层，以及位于两者之间的多个隐藏层。网络越深，隐藏层就越多。网络参数表示层与层之间连接的强度。训练网络的过程会确定参数的取值。训练完成后，人工神经网络便可作为一种模型，将输入（例如图像）转换为输出（标签或类别）。

前馈神经网络算法的用途十分广泛，相关的近期文献数量可观且内容多样，尤其涉及应用方面。例如，参见 [ASA+23，MM23，KG22，CC21，MCA20，Han16]。

以下引理是再生核希尔伯特空间理论中的一个基本结果。详情可参见 [Sza83，Sza15a，Sza15b，Sza21，Sza04]，以及 [JT22] 和其中引用的论文。

引理 2.1。取一个正定核 $X \times X \xrightarrow{K} \mathbb{R}$(或 $\mathbb{C}$)，令 $\mathscr{H}_K$ 表示相应的再生核希尔伯特空间。那么，函数 F 在 X 上属于 $\mathscr{H}_K$，当且仅当存在常数 $C_F < \infty$，使得对所有 $n \in \mathbb{N}$、所有 $(\xi_i)_{i=1}^n$、$\xi_i \in \mathbb{R}$(或 $\mathbb{C}$)，以及所有 $(x_i)_{i=1}^n$、$x_i \in X$，以下估计成立：

$$\left|\sum_{i=1}^{n} \xi_i F(x_i)\right|^2 \leq C_F \sum_{i=1}^{n}\sum_{j=1}^{n} \overline{\xi}_i \xi_j K(x_i, x_j). \tag{2.1}$$

注记 2.2。采用构造 $K \mapsto \mathscr{H}_K$（指固定正定核 K 的再生核希尔伯特空间），我们得到以下两个结论：

(1) 对所有 $x \in X$，函数 $K_x := K(\cdot, x)$ 属于 $\mathscr{H}_K$；(2) 对所有 $F \in \mathscr{H}_K$ 和 $x \in X$，有

$$F(x) = \langle F, K(\cdot, x)\rangle_{\mathscr{H}_K}, \tag{2.2}$$

也就是说，函数 F 在 $\mathscr{H}_K$ 中的值可通过内积 $\langle\cdot,\cdot\rangle_{\mathscr{H}_K}$ 和核函数再现。

除（2.2）之外，我们还将讨论相对再生核以及相对再生核希尔伯特空间。如 [AJV14]，所述，相对再生性质可表示为以下形式

$$F(y) - F(x) = \langle F, v_{x,y}(\cdot)\rangle_{\mathscr{H}_{rel}}, \tag{2.3}$$

现在对所有点对 $x,\ y \in X$ 均成立。因此，这意味着存在双指标核函数 $v_{x,\ y} \in \mathscr{H}_{rel}$。

本文将讨论一类 $\mathscr{H}_{rel}$ 空间，见下文定理3.4 。其中的设定为 $X = \mathbb{R}$，而相对核函数$v_{a,\ b}$ 则作为前馈神经网络算法类别的激活函数，参见例如图3.1。

文献2.3在 [AJV14] 中对其进行了系统研究，其中证明，相对再生的设定由条件负定函数刻画。

现在回顾标准一维布朗运动对应的再生核希尔伯特空间。（例如参见 [JT22、AJ21、JT21、JT20]。）

引理 2.3。 当 K 是定义在 $\mathbb{R}_{\geq} \times \mathbb{R}_{\geq}$ 上的布朗运动核时，即

$$K(x,y) = x \wedge y = \frac{|x| + |y| - |x-y|}{2}, \quad x, y \in \mathbb{R}_{\geq}, \tag{2.4}$$

相应的再生核希尔伯特空间 $\mathscr{H}_K$ 是由定义在 $\mathbb{R}$ 上的绝对连续函数f构成的希尔伯特空间，这些函数的导数 $f' = df/dx$属于 $L^2(\mathbb{R})$，且 $f(0) = 0$。此外，

$$\|f\|^2_{\mathscr{H}_K} = \int_{\mathbb{R}_{\geq}} |f'(x)|^2 \, dx, \quad \textit{for all } f \in \mathscr{H}_K. \tag{2.5}$$

证明。 关键观察是，若 $x > 0$，函数

$$\mathbb{R}_{\geq} \ni y \longmapsto F_x(y) := K(y,x) = \begin{cases} y & \text{if } y \leq x \\ x & \text{if } y > x \end{cases} \tag{2.6}$$

具有弱导数。事实上，我们有

$$\frac{dF_x}{dy} = \chi_{[0,x]}, \tag{2.7}$$

即区间 $[0,x]$ 的示性函数。因此，若 f 是一个满足$f' \in L^2(\mathbb{R})$ 且 $f(0) = 0$ 的函数，则

$$f(x) = f(x) - f(0) = \int_0^x f'(y)\, dy = \int_{\mathbb{R}} F'_x(y)\, f'(y)\, dy, \tag{2.8}$$

并且，(2.8) 中的右端项是由 (2.5) 右端项定义的希尔伯特空间中的内积。

相应的推论可由再生核希尔伯特空间的一般理论得出。回想一下，核的再生核希尔伯特空间是函数的希尔伯特空间完备化

$$y \longmapsto K(x,y) \tag{2.9}$$

当 x 在 $\mathbb{R}$ 中变化时。如果 $K(x,\ y) = x \wedge y$，

$$\langle K(\cdot, x_1), K(\cdot, x_2) \rangle_{\mathscr{H}_K} = K(x_1, x_2) = x_1 \wedge x_2,$$

我们可以如下计算：

$$\begin{aligned} \int_{\mathbb{R}} \left(\frac{d}{dy} K(\cdot, x_1)\right) \left(\frac{d}{dy} K(\cdot, x_2)\right) dy &= \int_{\mathbb{R}} \chi_{[0,x_1]}(y)\, \chi_{[0,x_2]}(y)\, dy \\ &= \lambda([0,x_1] \cap [0,x_2]) \\ &= x_1 \wedge x_2 = K(x_1, x_2) \end{aligned}$$

其中 $\lambda = dy$ 表示勒贝格测度。 □

注记 2.4. 注意，满足以下条件的函数 $v_{a,b}$ （称为偶极子）中 $\mathscr{H}_K$

$$f(b)-f(a)=\langle f, v_{a,b}\rangle, \quad \text{for all } f\in\mathscr{H}_K$$

（见（2.5）和（2.8））：

$$v_{a,b}(x)=\begin{cases}0 & \text{if } x<a\\ x-a & \text{if } a\le x<b\\ b-a & \text{if } x>b,\end{cases}$$

如图 2.1所示。另请参见定理3.4和图 3.1，以及第 4节中的迭代。

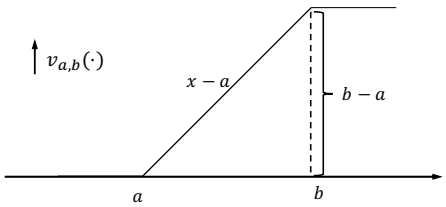


图 2.1。 生成偶极函数 $\{v_{a,b}\}$按成对索引 a， b 使得 $a<b$。请与下方的图 3.1 比较。

诱导度量

对于 $X\times X$ 上的一般正定核 K，在 X 上可诱导出一个度量，

$$d_K : X\times X\to\mathbb{R}_+$$

定义为（例如参见 [AJ21]）

$$d_K(x,y)=\|K(\cdot,x)-K(\cdot,y)\|^2_{\mathscr{H}_K}. \tag{2.10}$$

特别地，

$$d_K(x,y)=K(x,x)+K(y,y)-2\Re\{K(x,y)\}.$$

注意， $d_K^{1/2}$ 也是 $X\times X$ 上的一个度量。

例 2.5。 对于 $K(x,\ y)=x\wedge y$ 在 $\mathbb{R}\times\mathbb{R}$ 如（2.4）中所示，

$$\begin{aligned}\|K(\cdot,s)-K(\cdot,t)\|^2_{\mathscr{H}_K} &= \left\|(\cdot\wedge s)'-(\cdot\wedge t)'\right\|^2_{L^2}\\ &= \left\|\chi_{[0,s]}-\chi_{[0,t]}\right\|^2_{L^2}\\ &= |s-t|.\end{aligned}$$

以下结果讨论一类一般框架，其中包含成对的集合 X 和 Y，每个集合都配备一个正定核， K 分别为 L， K 定义在 X上，以及 L 定义在 Y上。为了实现前馈神经网络函数，我们将提出一个显式框架（见（2.11）和（2.20）），使我们能够将（非线性）函数f： $X\to Y$ 转化为在相应再生核希尔伯特空间之间作用的线性算子 T_f 和 $\mathscr{H}_K$。这是一种表示方式，即函数的复合会映射为相应线性算子的乘积。需要注意的是，线性算子 T_f 通常可能是无界的。尽管如此，我们将证明这些算子仍属于一类具有谱分解的算子，见定理 2.11和推论2.12。

定理 2.6。考虑正定核 $X \times X \xrightarrow{K} \mathbb{R}$ 和 $Y \times Y \xrightarrow{L} \mathbb{R}$。设$f\colon X \to Y$关于诱导度量 d_K、d_L是利普希茨连续的，即

$$d_L\left(f\left(x\right), f\left(y\right)\right) \leq c_f d_K\left(x, y\right), \quad x, y \in X,$$

对于某个常数 c_f。定义算子 $T_f\colon \ \mathscr{H}_K \to \mathscr{H}_L$ 为

$$T_f\left(K_x\right)\left(y\right) = L\left(f\left(x\right), y\right)$$

并利用线性性和稠密性将其延拓。

那么，对于任意固定的 $c \in Y$，函数

$$F_f : X \to \mathbb{R}, \quad F_f\left(x\right) := L\left(f\left(x\right), c\right) \tag{2.11}$$

属于再生核希尔伯特空间 $\mathscr{H}_K$ 当且仅当

$$L_c \in dom(T_f^*), \tag{2.12}$$

伴随算子的定义域。

此外，

$$\begin{gathered} F_{f,y}\left(x\right) := L\left(f\left(x\right), y\right) \in \mathscr{H}_K, \ \forall y \in Y \\ \Updownarrow \\ T_f \text{ is closable.} \end{gathered}$$

（另见定理2.8。）

证明。设定理陈述中的条件成立，即，$X \xrightarrow{f} Y$关于两个度量连续，d_K 在 X 上且 d_L 在 Y 上；因此，特别地，对于点对 $x_1, x_2 \in X$，有

$$\begin{aligned} d_K\left(x_1, x_2\right) &= \left\|K\left(\cdot, x_1\right) - K\left(\cdot, x_2\right)\right\|_{\mathscr{H}_K}^2 && (2.13) \\ &= K\left(x_1, x_1\right) + K\left(x_2, x_2\right) - 2K\left(x_1, x_2\right). && (2.14) \end{aligned}$$

我们还固定一个点 $c \in Y$，并设 $F = F_{f,c}$，定义如下：

$$F\left(x\right) = L\left(f\left(x\right), c\right), \text{ for all } x \in X,$$

因此 $F\colon \ X \to \mathbb{R}_+ \cup \{0\}$。

现在，对于每个 N，以及每个子集 $S_N = (x_1, x_2, \ldots\ldots, x_N) \subset X$，考虑以下矩阵运算（在 N 维空间中）：

$$\underbrace{F\big|_N := \begin{bmatrix} F\left(x_1\right) \\ \vdots \\ F\left(x_N\right) \end{bmatrix}}_{\text{列向量}}, \text{ and } \quad \underbrace{Q_N := \left|F\right|_N\rangle\langle F\big|_N\Big|}_{\text{matrix of a rank-1 operator}}, \tag{2.15}$$

即，作用于 $\mathbb{R}^N$ 上的秩一算子，用狄拉克记号表示，定义为

$$Q_N\left(\xi\right) = \left\langle F_N, \xi\right\rangle F_N \tag{2.16}$$

对所有 $\xi \in \mathbb{R}^N$ 均成立。令

$$K_N := \left(K\left(x_i, x_j\right)\right)_{i,j=1}^N = \begin{bmatrix} K\left(x_1, x_1\right) & \cdots & K\left(x_1, x_N\right) \\ \vdots & \ddots & \vdots \\ K\left(x_N, x_1\right) & \cdots & K\left(x_N, x_N\right) \end{bmatrix}, \tag{2.17}$$

一个样本矩阵。

对于所有正定 $N \times N$ 矩阵构成的凸锥，我们引入以下熟知的序关系：当且仅当（定义）存在 $\exists C < \infty$ 使得

$$\xi^T K_N \xi \leq C \xi^T K'_N \xi \text{ for all } \xi \in \mathbb{R}^N. \tag{2.18}$$

现在应用引理2.1可知，定理中的断言等价于存在一个有限常数 C (与 $S_N = (x_i)_{i=1}^N$ 无关 $Q_N \ll_C K_N$，使得估计

$$\left| \sum_i \xi L\left(f\left(x_i\right), c\right) \right|^2 \leq C \sum_i \sum_j \xi_i K\left(x_i, x_j\right) \xi_j \tag{2.19}$$

对所有 N、所有 $S_N = \{x_i\}_{i=1}^N$ 和所有 $\xi \in \mathbb{R}^N$ 均成立。根据定理中关于 f 的假设可得此结论。详情见下文。□

定理2.6的总结： 从 K 在 $X \times X$上的正定性， L 在 $Y \times Y$上的正定性，以及 f : $X \to Y$开始。我们在 d_K 上引入度量 X，并在 d_L 上引入度量 Y，然后考察f 的连续性或利普希茨性。为了得到所需结论

$$\left(x \longmapsto L\left(f\left(x\right), y\right)\right) \in \mathscr{H}_K,$$

我们必须引入一个算子 $T_f : \mathscr{H}_K \to \mathscr{H}_L$。恰当的选择是

$$L_y \in dom(T_f^*).$$

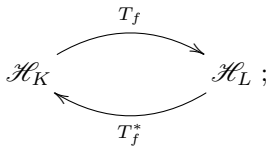
详情如下:

固定两个核 K 和 L，假设它们分别在 $X \times X$和 $Y \times Y$上是正定的。转到相应的再生核希尔伯特空间 $\mathscr{H}_K$ 和 $\mathscr{H}_L$。

问题。 求函数 $X \xrightarrow{f} Y$ 所满足的条件，使得对于$\forall y \in Y$，诱导函数

$$\underbrace{\left(X \ni x \longmapsto L\left(f\left(x\right), y\right)\right)}_{F_{f,y}(\cdot) \text{ as a function on } X} \in \mathscr{H}_K. \tag{2.20}$$

下面强调的论证通过对偶算子（有界）进行



但无界情形也很有意思。

关于在不对 $X \xrightarrow{f} Y$ 加任何附加假设的情况下算子 $T_f : \mathscr{H}_K \to \mathscr{H}_L$ 的定义的一些说明。

我们定义

$$T_f\left(K_x\right)(y) = L\left(f\left(x\right), y\right);$$

因此，我们将 T_f 扩展到线性组合:

$$\mathscr{D}_K := \{ \underbrace{\sum_i c_i K_{x_i}}_{\text{function on } X} \} \xrightarrow{T_f} \underbrace{\sum_i c_i L\left(f\left(x_i\right), \cdot\right)}_{\text{function on } Y}. \tag{2.21}$$

但是，为了使（2.21）有意义且定义良好，我们必须谨慎处理等价类。

如果 $f:\ X \to Y$ 是一个一般函数，则（广义）算子（2.21）可能不可闭。不过，我们仍然可以定义其伴随算子 T_f^*，但其定义域可能很“小”。

Set

$$\mathscr{D}_K := span\left\{K_x\right\}_{x\in X}, \tag{2.22}$$

那么（定义）向量 $\psi \in \mathscr{H}_L$ 属于 $dom(T_f^*)$ 当且仅当存在 $\exists C_\psi < \infty$ 使得

$$\left|\left\langle T_f\varphi, \psi\right\rangle_{\mathscr{H}_L}\right| \le C_\psi \left\|\varphi\right\|_{\mathscr{H}_K}, \quad \forall \varphi \in \mathscr{D}_K. \tag{2.23}$$

然后我们令 $T_f^*\psi =$ 为该方程的解

$$\left\langle T_f\varphi, \psi\right\rangle_{\mathscr{H}_L} = \left\langle \varphi, T_f^*\psi\right\rangle_{\mathscr{H}_K}, \quad \forall \varphi \in \mathscr{D}_K. \tag{2.24}$$

令 K，L，f满足所述条件，并假设对于某个 $y \in Y$，有 $L_y \in dom\left(T_f^*\right)$，那么（2.24）对于 $L_y \in \mathscr{H}_L \xrightarrow{T_f^*} \mathscr{H}_K$，$\varphi = K_x$，$\psi = L_y$给出

$$x \longmapsto T_f^*\left(L_y\right)(x) = L\left(f\left(x\right), y\right) \in \mathscr{H}_K.$$

因此，定理 2.6 中的结论

$$\left(x \longmapsto L\left(f\left(x\right), y\right)\right) \in \mathscr{H}_K \tag{2.25}$$

成立当且仅当 $L_y \in dom(T_f^*)$。

在这种情况下，(2.21) 不存在任何困难，因此我们得到一个对偶对 T_f 和T_f^*，

$$\left\langle T_f\varphi, \psi\right\rangle_{\mathscr{H}_L} = \left\langle \varphi, T_f^*\psi\right\rangle_{\mathscr{H}_K} \tag{2.26}$$

$\overset{T_f}{\frown}$对于 $\forall \varphi \in \mathscr{D}_K$，以及 $\psi \in \mathscr{D}_L$，

$\mathscr{H}_K \quad \mathscr{H}_L \circ \prec^{T_f^*}$ $\quad$

令 $\varphi = K_x$，并令 $\psi = L_y$，(2.26) 推出

$$T_f^*\left(L_y\right)(x) = L\left(f\left(x\right), y\right) = \left(T_f\left(K_x\right)\right)(y). \tag{2.27}$$

但前一个条件 $L_y \in dom(T_f^*)$ (参见 (2.26)) 等价于断言 $T_f^*\left(L_y\right) \in \mathscr{H}_K$，根据 (2.27)，这便是定理 2.6的结论。

推论 2.7（复合算子）。设 X，Y，K 和 L 如上所述；特别地，K 是定义在 $X \times X$上的一个固定正定核，而再生核希尔伯特空间 $\mathscr{H}_K$ 是定义在 X上的标量值函数的希尔伯特空间。类似地，$\mathscr{H}_L$ 是定义在 Y上的标量值函数的希尔伯特空间。$\mathscr{H}_K$ 和 $\mathscr{H}_L$ 都满足再生核希尔伯特空间的定义公理；见上文引理 *2.1* 。如前所述，每个函数 f，$X \xrightarrow{f} Y$，都会诱导一个线性算子

$$\mathscr{H}_K \xrightarrow{T_f} \mathscr{H}_L, \tag{2.28}$$

其定义域稠密于 $\mathscr{D}_K$；见定理 *2.6*的陈述。对于伴随算子T_f^*，

$$\mathscr{H}_L \xrightarrow{T_f^*} \mathscr{H}_K, \tag{2.29}$$

我们有如下结论：对于函数 ψ 在 $\mathscr{H}_L$中，以下两种刻画以及（*2.31*）成立：*(2.30)*

$$\psi \in dom(T_f^*) \tag{2.30}$$

$$\Updownarrow$$
$$\psi \circ f \in \mathscr{H}_K. \tag{2.31}$$

若肯定，

$$T_f^*(\psi) = \psi \circ f : X \to \mathbb{R}, \tag{2.32}$$

即， T_f^* 是复合算子。

证明。 (2.30)⇒(2.31)。假设 (2.30)，然后应用 (2.23)，得到 $C_\psi < \infty$ 其中：

$$\left| \sum_i c_i \left(T_f^*(\psi) \right)(x_i) \right|^2 \le C_\psi \sum_i \sum_j c_i c_j K(x_i, x_j). \tag{2.33}$$

But

$$\begin{aligned} T_f^*(\psi)(x_i) &= \left\langle K_{x_i}, T_f^*(\psi) \right\rangle_{\mathscr{H}_K} \\ &= \left\langle T_f(K_{x_i}), \psi \right\rangle_{\mathscr{H}_L} \\ &= \left\langle L(f(x_i), \cdot), \psi \right\rangle_{\mathscr{H}_L} \\ &= \psi(f(x_i)), \end{aligned} \tag{2.34}$$

在最后一步中，我们使用了 $\mathscr{H}_L$ 的再生核希尔伯特空间性质。代入 (2.33) 得

$$\left| \sum_i c_i \psi(f(x_i)) \right|^2 \le C_\psi \sum_i \sum_j c_i c_j K(x_i, x_j), \tag{2.35}$$

因此，由于引理 2.1 应用于 $F = \psi \circ f$，结论 (2.31) 得证。

(2.31)⇒(2.30)。假设 (2.31) 成立，则反过来推导上述论证可得

$$\left| \left\langle T_f \underbrace{\left(\sum_i c_i K_{x_i} \right)}_{\in \mathscr{D}_K}, \psi \right\rangle_{\mathscr{H}_L} \right| \le \sqrt{C_\psi} \left\| \sum_i c_i K_{x_i} \right\|_{\mathscr{H}_K} \tag{2.36}$$

这表明 $\psi \in dom(T_f^*)$，这就是条件 (2.30)。现在将其与 (2.34) 结合，可得 (2.32) 对 ψ 成立，即 $T_f^* \psi = \psi \circ f$ 成立。 □

2.1. **基底方法**。令 X, Y, K, L, f 如常定义，并定义 T_f: $\mathscr{H}_K \to \mathscr{H}_L$。由于 K 在 $X \times X$ 上正定， 再生核希尔伯特空间 $\mathscr{H}_K$ 允许存在标准正交基 $\{h_i\}_{i \in \mathbb{N}}$, $h_i \in \mathscr{H}_K$；由一般理论可得逐点公式：

$$K(x_1, x_2) = \sum_{i \in \mathbb{N}} h_i(x_1) h_i(x_2). \tag{2.37}$$

那么，定理 2.6 中的条件等价于以下条件：

$$(X \ni x \longmapsto L(f(x), y)) \in \mathscr{H}_K \tag{a}$$

$$\Updownarrow$$

$$\sum_{i \in \mathbb{N}} \left| \left(T_f(h_i) \right)(y) \right|^2 < \infty. \tag{b}$$

证明。 $(a) \Rightarrow (b)$ 的详细过程如下；但逆向推论也可通过相同的论证得到。因此，由 (a) 可知， $L_y \in dom(T_f^*)$，从而：

$$T_f^*(L_y)(\cdot) \in \mathscr{H}_K. \tag{2.38}$$

由于 $\{h_i\}$ 是 $\mathscr{H}_K$ 中的标准正交基,

$$\sum_i \left|\left\langle h_i, \underbrace{T_f^*\left(L_y\right)}_{\in \mathscr{H}_K}\right\rangle\right|^2 = \left\|T_f^*\left(L_y\right)\right\|_{\mathscr{H}_K}^2 < \infty. \tag{2.39}$$

但从等式左侧(2.39):

$$\left\langle h_i, T_f^*\left(L_y\right)\right\rangle_{\mathscr{H}_K} = \left\langle T_f\left(h_i\right), L_y\right\rangle_{\mathscr{H}_L} = \left(T_f\left(h_i\right)\right)(y),$$

并且 (b) 成立。 □

关键问题：何时$F_{f,\ y}(\cdot) \in \mathscr{H}_K$**?** 关于哪些函数 $X \xrightarrow{f} Y$ 具有以下性质这一问题，最简洁的回答是

$$F_{f,y}(x) = L\left(f(x), y\right) \text{ is in } \mathscr{H}_K \tag{2.40}$$

是这样的:

定理 2.8。 设给定 K, L 和 f，则

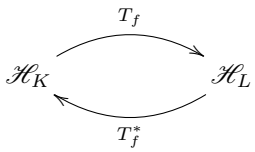
$$F_{f,y} \text{ in } (2.40) \text{ is in } \mathscr{H}_K \Longleftrightarrow L_y \in dom(T_f^*), \tag{2.41}$$

其中算子 T_f： $\mathscr{H}_K \to \mathscr{H}_L$ 定义为

$$T_f\left(K_x\right) := L\left(f(x), \cdot\right).$$

此外，对于所有 *2.41*， $y \in Y \Longleftrightarrow T_f$ 都是可闭的。

2.2. **算子的对偶对**。 考虑一对定义在稠密定义域上的对称算子:



($T = T_f$，因为它将取决于 f)，其中

$$span\left\{K_x\right\}_{x \in X} \text{ is dense in } \mathscr{H}_K \tag{2.42}$$

and

$$span\left\{L_y\right\}_{y \in Y} \text{ is dense in } \mathscr{H}_L \tag{2.43}$$

使得

$$K_x \in dom(T_f), \text{ and} \tag{2.44}$$
$$L_y \in dom(T_f^*) \tag{2.45}$$

其中“dom”分别表示相应的算子定义域。

注。 我们注意到

$$T_f\left(K_x\right)(\cdot) = L\left(f(x), \cdot\right) \in \mathscr{H}_L$$

总是定义良好，且定义域稠密，但关键在于 T_f^*.

另请注意，(2.45) 是定理 2.6中的条件。

设 $f: X \to Y$ 如前所述，并固定两个正定核 K 和 L 。接下来，我们通过如下定义引入相应的（稠密定义）算子 $T_f: \mathscr{H}_K \to \mathscr{H}_L$

$$T_f(K_x) = L(f(x), \cdot) \in \mathscr{H}_L. \tag{2.46}$$

记号与约定。 $K_x(\cdot)$ 是 $\mathscr{H}_K$ 中通常所说的核函数:

$$K_x(t) = K(x,t), \ \forall t \in X \text{ and similarly,} \tag{2.47}$$

$$L_y(u) = L(y,u), \ \forall u \in Y. \tag{2.48}$$

引理 2.9。当 $L_y \in dom(T_f^*)$ 则

$$\left(T_f^*(L_y)\right)(x) = L(f(x), y) \ \ on\ X, \tag{2.49}$$

等价地,

$$T_f^*(L_y)(\cdot) = L(f(\cdot), y) \ \ on\ X. \tag{2.50}$$

证明（*2.49*）。结论（2.49）等价于以下断言:

$$\left\langle \underbrace{T_f(K_x)(\cdot)}_{L(f(x),\cdot)}, L_y \right\rangle_{\mathscr{H}_L} = \left\langle K_x, \overbrace{\underbrace{L(f(\cdot), y)}_{T_f^*(L_y)}}^{\in \mathscr{H}_K} \right\rangle_{\mathscr{H}_K}$$

$$\searrow \qquad \swarrow$$

$$L(f(x), y)$$

结论（2.49）成立，因为相应的核函数张成稠密子空间。□

回顾,

$$\begin{array}{c} \text{the function } x \longmapsto F_{f,y}(x) = L(f(x), y) \in \mathscr{H}_K \\ \Updownarrow \\ L_y \in dom(T_f^*). \end{array}$$

假设 $L_y \in dom(T_f^*)$，然后将函数在 $\mathscr{H}_K$ 至$F_{f,\ y}(\cdot)$ 中的条件应用于此，因此 $\forall n$, $\forall (x_i)_1^n, \forall (c_i)_1^n, c_i \in \mathbb{R}$:

$$\begin{aligned} \left| \sum_i c_i F_{f,y}(x_i) \right|^2 &= \left| \sum_i c_i L(f(x_i), y) \right|^2 \\ &\leq \left| \left\langle \sum_i c_i K_{x_i}, T_f^*(L_y) \right\rangle_{\mathscr{H}_K} \right|^2 \\ &\underset{\text{Schwarz}}{\leq} \left\| \sum_i c_i K_{x_i} \right\|_{\mathscr{H}_K}^2 \left\| T_f^*(L_y) \right\|_{\mathscr{H}_K}^2 \\ &= \sum_i \sum_j c_i c_j K(x_i, x_j) \left\| T_f^*(L_y) \right\|_{\mathscr{H}_K}^2. \end{aligned}$$

引理 2.10. 下面的蕴含关系双向成立:

$$X \ni x \longmapsto L\left(f\left(x\right), y\right) \in \mathscr{H}_K \text{ for } \forall y$$
$$\Updownarrow$$
$$\text{the condition in (2.6) is satisfied.}$$

即使我们固定了 $y \in Y$，那么

$$L_y \in dom(T_f^*) \Longleftrightarrow \left(x \longmapsto L\left(f\left(x\right), y\right)\right) \in \mathscr{H}_K. \tag{2.51}$$

证明概要。 根据定义， $L_y \in dom(T_f^*)$， $\exists C_y < \infty \Longleftrightarrow$

$$\left|\left\langle T_f\varphi, L_y\right\rangle_{\mathscr{H}_L}\right| = \left|\left(T_f\left(\varphi\right)\right)\left(y\right)\right| \leq C_y \left\|\varphi\right\|_{\mathscr{H}_K}$$

对 $\forall \varphi \in span\left\{K_x\right\}_{x\in X}$. 成立，但是

$$T_f\left(K_x\right)\left(y\right) = L\left(f\left(x\right), y\right), \text{ and} \tag{2.52}$$

$$\left|\sum_i c_i \underbrace{L\left(f\left(x_i\right), y\right)}_{F_{f,y}(x_i)}\right|^2 = \left|\left\langle \sum_i c_i K_{x_i}, T_f^*\left(L_y\right)\right\rangle_{\mathscr{H}_K}\right|^2 \tag{2.53}$$

$$\leq \left\|T_f^*\left(L_y\right)\right\|_{\mathscr{H}_K}^2 \underbrace{\sum_i \sum_j c_i c_j K\left(x_i, x_j\right)}_{<\infty} \tag{2.54}$$

因此，根据 $\mathscr{H}_K$ 的基本引理（参见定理2.6的证明），我们得出函数 $F_{f,y} \in \mathscr{H}_K$，即 $\left(x \longmapsto L(f(x) \quad y)\right) \in \mathscr{H}_K$。

结论： 双向蕴含 $\Longleftrightarrow$ 在（2.51）中成立。 □

2.3. 函数与算子。一般而言，若 $T: \mathscr{H}_1 \to \mathscr{H}_2$ 是一个定义在稠密定义域 $\mathscr{D} \subset \mathscr{H}_1$ 上的算子，其中 $\mathscr{H}_i$, $i = 1,2$ 是两个希尔伯特空间，我们知道 T 可闭 $\Longleftrightarrow$ T^* 稠密定义，当且仅当 $dom(T^*)$ 在 $\mathscr{H}_2$ 中稠密（例如参见[JT21]）。因此，我们将此应用于 $T = T_f$, $\mathscr{H}_1 = \mathscr{H}_K$, $\mathscr{H}_2 = \mathscr{H}_L$，定理 2.6 中的条件 $\Longleftrightarrow$ $L_y \in dom(T_f^*)$ $\forall y \in Y$成立。由于 $span\left\{L_y\right\}_{y\in Y}$在$\mathscr{H}_L$中稠密，定理 2.6 $\Longrightarrow$ T_f 可闭。

给定上述 K 和 L，引入

$$\mathscr{F}_{ub}\left(K, L\right) = \left\{f : T_f \text{ is closable}\right\}, \text{ and} \tag{2.55}$$
$$\mathscr{F}_{b}\left(K, L\right) = \left\{f : T_f \text{ is bounded from } \mathscr{H}_K \text{ into } \mathscr{H}_L\right\}. \tag{2.56}$$

在这两种情况下，算子 $T = T_f$ 取决于函数 $X \xrightarrow{f} Y$的选择，但两个条件 (2.55) 和 (2.56) 并不相同:

$$\left(T_f\left(K_x\right)\right)\left(y\right) = L\left(f\left(x\right), y\right) = \left(\left(T_f\right)^*\left(L_y\right)\right)\left(x\right), \tag{2.57}$$

对所有 $x \in X$ 和 $y \in Y$ 均成立。详情见下文:

关于算子 T_f： $\mathscr{H}_K \to \mathscr{H}_L$。与前面一样， K 和L 是固定的正定核，而 $f: X \to Y$ 是一个函数。我们需要理解定理2.6的结论，即何时

$$\left(X \ni x \longmapsto L\left(f\left(x\right), y\right)\right) \in \mathscr{H}_K \text{ for all } y \in Y? \tag{2.58}$$

答案： （2.58） $\Longleftrightarrow$ $L_y \in dom\ (T_f^*)$ 。

注意，此时（2.58）中的函数为 T_f^* （L_y）；参见（2.57）。但请注意，从一个函数 $X \xrightarrow{f} Y$出发，要使（2.57）定义出一个良定义的线性算子 T_f，其定义域在 $\mathscr{H}_K$中稠密，需满足

$$T_f(K_x)(\cdot) = L(f(x), \cdot). \tag{2.59}$$

当 T_f有界时，情况很简单，因为此时 $dom(T_f^*) = \mathscr{H}_L$。在记号上，$L(f(x) \cdot) \in L_{f(x)} \in \mathscr{H}_L$，但我们还必须验证所有有限和对应的隐式核函数：

$$\sum_i \sum_j c_i c_j K(x_i, x_j) = 0 \Longrightarrow \sum_i \sum_j c_i c_j L(f(x_i), f(x_j)) = 0.$$

2.4. **当 $K = L$时**。如下面第 3节所示，在多层神经网络的应用中，当每一层都使用相同的正定核K 时，递归构造会简化。因此，下面我们专门讨论 $X = Y$，且 $K = L$的情形；参见定理 2.6 和 2.8中的设定。

定理 2.11。考虑定义在 K 上的正定核 $X \times X$，以及相应的再生核希尔伯特空间 $\mathscr{H}_K$，即 $\{K_x\}_{x \in X}$ 的希尔伯特完备化，其中 $K_x := K(\cdot, x)$。固定一个函数 $X \xrightarrow{f} X$，使其满足以下性质（见定理 *2.6*）：

$$(X \ni x \longmapsto K(f(x), y)) \in \mathscr{H}_K \text{ for all } y \in X. \tag{2.60}$$

因此，由下式定义的算子 $T_f : \mathscr{H}_K \to \mathscr{H}_K$

$$T_f(K(\cdot, y)) := K(f(\cdot), y) \tag{2.61}$$

是一个从 $\mathscr{H}_K$ 到 $\mathscr{H}_K$的稠密定义算子，其定义域为

$$\mathscr{D}_K := span\{K_x\}_{x \in X}. \tag{2.62}$$

(1) 那么，T_f 的闭包（也记作 T_f）定义良好且为正规算子，即两个算子 T_f 和 T_f^* 交换。

(2) 特别地，T_f 具有投影值谱分解，即存在定义在 $\mathscr{B}_{\mathbb{C}}$ (= 中的博雷尔子集) 上的投影值测度 $Q(\cdot)$，使得

$$T_f = \int_{spect(T_f)} \lambda Q(d\lambda) : \mathscr{D}_K \to \mathscr{H}_K. \tag{2.63}$$

证明。注意，第（2）部分由（1）以及正规算子的谱定理得出（在希尔伯特空间 $\mathscr{H}_K$中）。

部分（1）。引入算子 T_f^* 后，我们得到以下交换关系：

$$\begin{array}{ccc} K(\cdot, x) & \xrightarrow[T_f]{} & K(\cdot, f(x)) \in \mathscr{D}_K \\ {\scriptstyle T_f^*}\downarrow & & \downarrow{\scriptstyle T_f^*} \\ K(f(\cdot), x) & \xrightarrow[T_f]{} & K(f(\cdot), f(x)) \in \mathscr{H}_K \end{array}$$

图 2.2. T_f与 T_f^*的交换性。

这就是所需的结论（1）。 □

给定一个函数 $f:\ X\to X$，如定理 2.11所述。下面我们利用定理 2.11中对应的投影值测度 $Q^{(f)}$，从而建立从点对$(x,\ y)$（其中各点属于 X）到复测度系统的映射 $\mu_{(x,y)}$，这些复测度定义在 T_f的谱上。在该映射中，函数 f 的 n次复合迭代给出每个测度的第 n阶矩 $\mu_{(x,y)}$。

推论 2.12。设 K，X 和 f 如定理 *2.11*所述，并设 $Q=Q^{(K,f)}(\cdot)$ 为（*2.63*）中的相应投影值测度。则对于任意一对 $x,\ y\in X$，可得到相应的博雷尔测度

$$\mu_{x,y}^{(f)}(B)=\left\langle K_x,Q(B)K_y\right\rangle_{\mathscr{H}_K}=\left(Q(B)(K_y)\right)(x),\tag{2.64}$$

对所有 $B\in\mathscr{B}_{\mathbb{C}}$ 均成立。归纳地，令

$$f^{\circ n}=\underbrace{f\circ\cdots\circ f}_{n\ fold}$$

我们便得到相应复测度的以下矩公式：

$$\mu_{f^{\circ n}(x),y}^{(f)}(B)=\int_B\lambda^n\,\mu_{x,y}^{(f)}(d\lambda).\tag{2.65}$$

现在我们来讨论乘子 在再生核希尔伯特空间 $\mathscr{H}_K$中的作用。

定义 2.13. 若标量值函数 φ 定义在 X 上，且满足以下两个等价条件之一，则称其为 $\mathscr{H}_K$ 的乘子：

(1) 乘法算子 M_φ 通过 $M_\varphi F=\varphi F$（逐点乘积）作用于 $\mathscr{H}_K$，使 $\mathscr{H}_K$ 保持不变。(2) 伴随算子满足以下恒等式：

$$M_\varphi^*(K_x)=\varphi(x)K_x\ \text{for all}\ x\in X\tag{2.66}$$

其中 K_x 表示核函数 $K_x=K(\cdot,x)$。

注 2.14. （1）与（2）的等价性可参见再生核希尔伯特空间的标准参考文献，例如 [JT21]。

定理 2.15. 设 K 是定义在 $X\times X$上的一个固定正定核，且设 $\mathscr{H}_K$ 是相应的再生核希尔伯特空间。设 $X\xrightarrow{f}X$ 是一个满足（*2.60*）的函数，即（$X\ni x\longmapsto K$（f（x） y）$\in\mathscr{H}_K$ 对所有 $y\in X$成立。

那么，对于每个属于 $\mathscr{H}_K$ 的乘子 φ，我们有：

$$M_{(\varphi\circ f)}T_f^*=T_f^*M_\varphi.\tag{2.67}$$

证明。显然，结论（2.67）等价于以下形式：

$$T_fM_{(\varphi\circ f)}^*=M_\varphi^*T_f;\tag{2.68}$$

下面我们将证明（2.68）。

设 f 和 φ 如定理中所述。于是我们得到以下交换图：

$$\begin{array}{ccc} K(\cdot,x) & \xrightarrow[M^*_{(\varphi\circ f)}]{} & \varphi(f(x))\,K(\cdot,x) \\ {\scriptstyle T_f}\downarrow & & \downarrow{\scriptstyle T_f} \\ K(\cdot,f(x)) & \xrightarrow[M^*_{\varphi}]{} & \varphi(f(x))\,K(\cdot,f(x)) \end{array}$$

图 2.3. 与（2.68）对应的交换图。

在验证图 2.3中的断言时，我们用到了上文定理2.6 和 2.8 的结论。□

3. 神经网络——由正定核得到的激活函数

上一节中，我们介绍了正定核及其相关生成函数在神经网络算法中的应用。下面，我们将利用核分析来设计生成神经网络函数。

下一个定义采用文献中前馈函数的迭代生成方式，例如：[CC21、CKM22、Han16]。此处定义和引理3.2 中的递归步骤，应用了我们在上文定理2.6和2.8 所建立的一般框架。

定义 3.1。设 K 是定义在 $\mathbb{R}$ 上的一个正定核。具有核 K 的 l 层前馈网络是如下形式的函数

$$\begin{aligned} x \mapsto y_1 = K(A_1x+b_1,c_1) \mapsto y_2 = K(A_2y_1+b_2,c_2) \mapsto \cdots \\ \cdots \mapsto y_l = K(A_ly_{l-1}+b_l,c_l) \mapsto y_{out} = K(\langle a_{l+1},y_l\rangle + b_{l+1},c_{l+1}) \end{aligned}$$

其中

- $x \in \mathbb{R}^{n_0}$;
- $A_j \in \mathbb{R}^{n_j\times n_{j-1}}$, b_j, $c_j \in \mathbb{R}^{n_j}$ 对于 $j=1,\cdots,l$;
- $a_{l+1} \in \mathbb{R}^{n_l}$, b_{l+1}, $c_{l+1} \in \mathbb{R}$;

对于向量 $x,y \in \mathbb{R}^m$,

$$K(x,y) := [K(x_1,y_1),\cdots,K(x_m,y_m)].$$

引理 3.2. 设 $K(x,y) = \min(x,y)$，且 a,b,c,d 为非零常数。则
(1) $K(ax+b,c) = aK\left(x,a^{-1}(c-b)\right)+b$; (2) $K(K(x,a)\ b) = K(x,K(a,b))$;
(3) $K(dK(ax+b,c)+e,f) = daK\left(x,K\left(a^{-1}(c-b)\ \ a^{-1}\left(d^{-1}(f-e)\right.\right.\right.$
$\left.\left.\left.-b\right)\right)\right)+db+e$.

证明。(1) $K(ax+b,c) = \begin{cases} ax+b & x < a^{-1}(c-b) \\ c & x > a^{-1}(c-b) \end{cases}$

(2) 假设 $a<b$，则

$$K(K(x,a),b) = \begin{cases} x & x<a \\ a & x \geq a \end{cases} = K(x,a).$$

情形 $a>b$ 类似。

(3) 由 (1)–(2) 可得:

$$\begin{aligned} & K\left(dK\left(ax+b,c\right)+e,f\right) \\ = \; & dK\left(K\left(ax+b,c\right),d^{-1}\left(f-e\right)\right)+e \\ = \; & dK\left(aK\left(x,a^{-1}\left(c-b\right)\right)+b,d^{-1}\left(f-e\right)\right)+e \\ = \; & d\left\{aK\left(K\left(x,a^{-1}\left(c-b\right)\right),a^{-1}\left(d^{-1}\left(f-e\right)-b\right)\right)+b\right\}+e \\ = \; & daK\left(K\left(x,a^{-1}\left(c-b\right)\right),a^{-1}\left(d^{-1}\left(f-e\right)-b\right)\right)+db+e \\ = \; & daK\left(x,K\left(a^{-1}\left(c-b\right),a^{-1}\left(d^{-1}\left(f-e\right)-b\right)\right)\right)+db+e \end{aligned}$$

□

下文中，所有网络均限定为定义在 $\mathbb{R}^d$ 中的紧子集 Ω 上，例如 $\Omega = [0,1]^d$（超立方体）。这是出于神经网络训练中标准化处理的考虑。

在定理 3.3 和 3.4中，我们详细介绍了特定的相对再生核希尔伯特空间，其各自的偶极子系统（见（3.5））正是此处图 2.1 和 3.1中所示的广义 ReLU 函数。

这里我们指定由 $\mathbb{R}$ 索引的 K_1 布朗运动 W 的核。因此，$\mathbb{R}\times\mathbb{R}$ 上相应的正定核如下:

$$K_1\left(x,y\right)=\begin{cases}\left|x\right|\wedge\left|y\right|=\min\left(\left|x\right|,\left|y\right|\right) & \text{if } xy\geq 0 \text{ (so same sign)}\\ 0 & \text{if } xy<0, \text{ so opposite sign.}\end{cases} \tag{3.1}$$

证明。 核 K_1 与布朗运动 $\{W_x\}_{x\in\mathbb{R}}$之间的联系如下:

$$K_1\left(x,y\right)=\mathbb{E}\left(\left(W_x-W_0\right)\left(W_y-W_0\right)\right) \tag{3.2}$$

对所有 x，$y\in\mathbb{R}$。由此结合布朗运动增量独立性即可得到所断言的公式（3.1）。□

定理 3.3。 设 K_1 为 *3.1*）在 $\mathbb{R}\times\mathbb{R}$上的正定核，其对应的再生核希尔伯特空间为

$$\mathscr{H}_{K_1}=\left\{f:f'\in L^2\right\},\quad \left\|f\right\|^2_{\mathscr{H}_{K_1}}=\int\left|f'\right|^2d\lambda_1.$$

在 $\Omega=[0,1]^d$上，考虑正定核

$$K_d\left(x,y\right)=K_1\left(x_1,y_1\right)\cdots K_1\left(x_d,y_d\right),$$

因此

$$\mathscr{H}_{K_d}=\left\{f:\nabla f\in L^2\right\},\quad \left\|f\right\|^2_{\mathscr{H}_{K_d}}=\int\left|\nabla f\right|^2d\lambda_d,$$

其中 λ_d 表示 d维勒贝格测度。

给定 $f:\Omega\to\mathbb{R}$，以及一个固定的 $c\in\mathbb{R}$，令

$$F:\mathbb{R}^d\to\mathbb{R}^1,\quad F\left(x\right)=K_1\left(f\left(x\right),c\right).$$

那么，

$$F\in\mathscr{H}_{K_d}\iff\iint_{f^{-1}([0,c])}\left|\nabla f\right|^2d\lambda_d<\infty.$$

定理 3.4。 设 μ 是定义在$(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$上的一个非原子 σ-有限测度，并考虑$(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ 上的斯蒂尔切斯测度 dF，使得

$$dF \ll \mu \tag{3.3}$$

(绝对连续)。 那么，正定核对应的相对再生核希尔伯特空间 $\mathscr{H}_\mu$

$$K_\mu(A, B) = \mu(A \cap B) \tag{3.4}$$

由满足以下条件的函数 F 组成

$$F(b) - F(a) = \left\langle v_{a,b}^{(\mu)}, F \right\rangle_{\mathscr{H}_\mu} \tag{3.5}$$

$$\int_{\mathbb{R}} \left| \frac{dF}{d\mu} \right|^2 d\mu < \infty \tag{3.6}$$

其中，相对核 $v_{a\,b}^{(\mu)}$，如下：

$$\frac{dv_{a,b}^{(\mu)}}{d\mu}(x) = \chi_{[a,b]}(x),$$

参见图*3.1*。

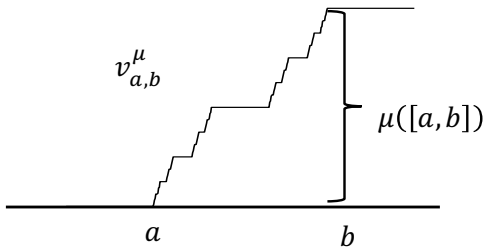


图 3.1。 偶极函数的示意图，它们能够重现该空间中函数值的差 $\mathscr{H}_\mu$。 与上文的图2.1 比较。

证明。 参见 [AJV14、 JT18]，详细内容见定理 3.3的证明。 □

注 3.5。 正定核 K_μ 是相对再生核希尔伯特空间 $\mathscr{H}_\mu$ 的“成因”，其定义域为 $\mathscr{B} \times \mathscr{B}$，其中 $\mathscr{B}$ 表示 σ的子集的博雷尔 $\mathbb{R}$-代数。利用 [JT18], 可验证

$$K_\mu(A, B) := \mu(A \cap B) \text{ for all } A, B \in \mathscr{B}. \tag{3.7}$$

我们还注意到， K_μ 是广义 μ-布朗运动$\{W_A^{(\mu)}\}_{A \in \mathscr{B}}$的协方差，即满足

$$\mathbb{E}\left(W_A^{(\mu)} W_B^{(\mu)}\right) = \mu(A \cap B) \text{ for all } A, B \in \mathscr{B}. \tag{3.8}$$

$W^{(\mu)}$ 对应的伊藤引理定义于$\mathbb{R}$ 上的可微函数 f，具体为

$$f\left(W_A^{(\mu)}\right) - f(0) = \int_A f'\left(W_t^{(\mu)}\right) dW_t^{(\mu)} + \frac{1}{2}\int_A f''\left(W_t^{(\mu)}\right) \mu(dt). \tag{3.9}$$

特别地，测度 μ 是 $W_t^{(\mu)}$ 的二次变差。

4. 分形图像的应用

近几十年来，人们逐渐发现，分形特征出现在各种数据集中，例如时间序列和图像分析中。（例如参见 [KLW20、KLLW21]。）或许，分形特征中最广为人知的例子包括精确的尺度对称性。通过预先规定的仿射映射系统，它们呈现为自相似性。其中一个特殊情形包括迭代函数系统（IFS）和最大熵测度，也称为IFS测度。较为熟悉的康托构造，例如按3或按4进行尺度变换，便是IFS测度的例子。对于这些情形中的每一种，我们在第3节提出的再生核希尔伯特空间框架，都是适用于此类神经网络算法的理想工具。特别地，这可以通过大量图像加以说明，例如生成5000幅图像，每幅图像都是二维或三维分形，带有随机旋转、缩放和彩色处理；其中一半的缩放因子为3，另一半的缩放因子为4。这将促成一个网络的训练，用于根据缩放因子对图像进行分类。

特别是，康托型激活函数，或类康托测度的累积分布函数（图 3.1），在 $[0,1]$ 的某些结构化子区间上导数为零。这一特性可能为神经网络带来多种益处。例如，此类函数可以在网络中引入稀疏性和正则化，从而提升其泛化性能并降低过拟合风险。此外，这些函数还能增强网络对输入数据中的噪声及其他扰动的鲁棒性，进而提升其在未见数据上的性能。此外，导数在子区间上为零的激活函数可以让网络学习数据中更复杂的非线性模式。这有助于提升网络的表达能力和灵活性，使其在更广泛的任务中更加准确有效。此外，由于激活函数的梯度具有良好的结构，此类函数还可以让网络更易于优化和训练，从而降低计算复杂度并提高训练算法的收敛速度。

更一般地，采用自定义激活函数的神经网络（例如，见图 2.1中的偶极子）使用带有可调参数的非标准激活函数，这些参数可在学习过程中进行训练和优化。这使网络能够学习输入数据与输出数据之间更复杂的非线性关系，从而提高网络预测的准确率。

在各种应用中，使用带有可训练参数的自定义激活函数都很有用，例如图像识别、自然语言处理和时间序列预测。它还可用于提升其他机器学习算法的性能，例如决策树和支持向量机（例如，参见 [CST01、PORSTS21、HSTHD11、HST10、RSSST06、STWCK05]）。

下面我们将自定义激活函数应用于卷积神经网络，对分形图像进行分类。在此设置中，激活函数的设计应能捕捉分形图像所特有的复杂自相似模式。网络使用带有相应标签的分形图像数据集进行训练，并采用基于梯度的算法（例如随机梯度下降）进行优化。训练完成后，网络将用于对新的分形图像进行分类，并以较高的准确率预测其所属类别。

在以下示例中，我们在 Mathematica 中生成了一个包含 15,000 张类康托尔三维图像的数据集[1] 。每张图像的参数（例如缩放因子、观察角度和缩放因子）均匀分布。图 4.1 展示了部分图像。

图像根据缩放因子分为三类，分别标记为类别“1”、“2”和“3”。整个数据集划分为训练集（大小为 = 10,000）、验证集（大小为 = 2,500）和测试集（大小为 =2,500）。任务是使用训练集训练一个三分类器，并利用验证集进行模型选择，随后在测试图像上测试其性能。

实验中，我们使用 Keras 实现了一个小型卷积神经网络。其架构如图 4.2所示。记录模型在 20 个训练轮次中的损失和准确率（图 4.3）。

与架构相同的标准 ReLU网络（图 4.3a）相比，使用类康托尔激活函数更有助于减少过拟合（图4.3b）；预计通过系统地进行超参数调优，此类网络有望在某些应用中优于 ReLU网络。

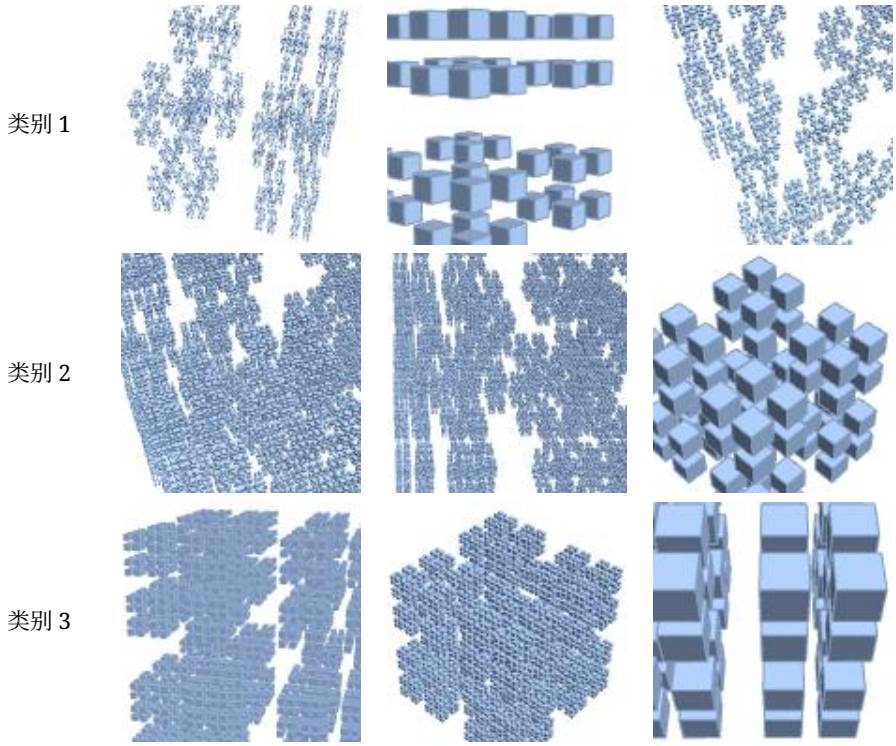


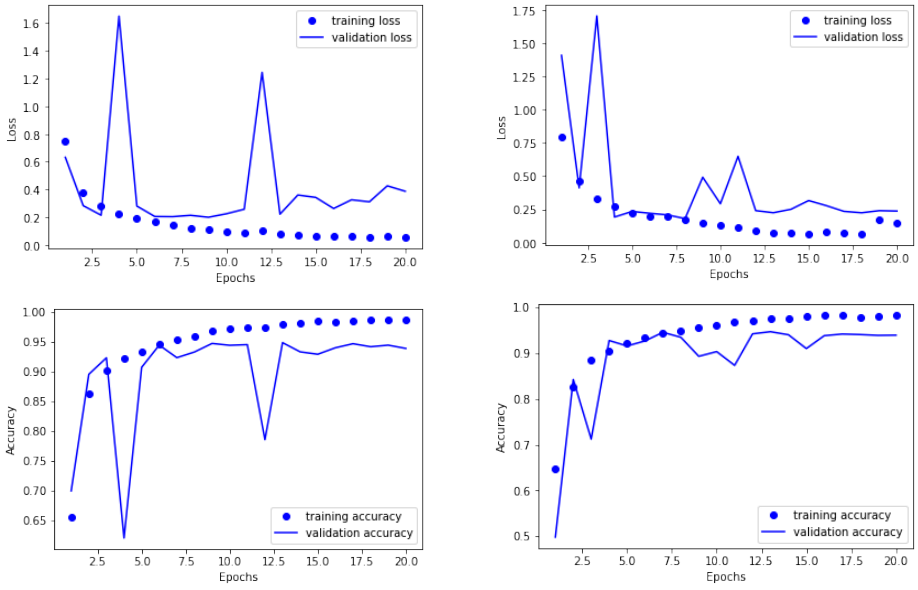
图4.1。三维康托尔图像数据集的一个随机样本。

[1]可在 https://www.kaggle.com/dsv/4791103 获取。

模型：“Cantor 激活函数”

层（类型）	输出形状	参数 #
输入_1 (输入层)	(None, 128, 128, 3)	0
重缩放（重缩放）	(None, 128, 128, 3)	0
二维卷积（二维卷积）	(None, 126, 126, 16)	448
最大_池化2d (最大池化2D)	(None, 63, 63, 16)	0
二维卷积_1 (二维卷积)	(None, 61, 61, 32)	4640
最大_池化2d_1 (最大池化2D)	(None, 30, 30, 32)	0
二维卷积_2 (二维卷积)	(None, 28, 28, 64)	18496
最大值_池化二维_2 (最大池化二维)	(None, 14, 14, 64)	0
二维卷积_3 (二维卷积)	(None, 12, 12, 128)	73856
展平（展平）	(None, 18432)	0
全连接层（全连接层）	(None, 3)	55299

图4.2。用于分形图像分类的 AConvNet。



(a) ReLU 激活 (b) 类康托尔激活函数

图4.3. 使用 AConvNet 展示的训练损失和验证损失。

参考文献

[AJ21] Daniel Alpay 和 Palle E. T. Jorgensen，再生核希尔伯特空间的新刻画及其在度量几何中的应用，Opuscula Math. 41 3, 283 2021 300 4302453 ( ), no. – . MR [AJV14] Daniel Alpay、Palle J orgensen 和 Dan Volok，相对再生核希尔伯特空间，Proc. Amer. Math. Soc.
142 2014 11, 3889 3895 3251728 ( ),
[23]
no. – . MR AK George A. Anastassiou 和 Seda Karateke，Richards 曲线诱导的巴拿赫空间值常微分方程及分数阶神经网络逼近，Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat. RACSAM 117 (2023), no. 1, Paper No. 14. MR 4505203[ASA+23] Mohd Rashid Admon、Norazak Senu、Ali Ahmadian、Zanariah Abdul Majid 和 So- heil Salahshour，一种基于前馈神经网络、用于求解分数阶微分方程的新型高效算法，Commun. Nonlinear Sci. Numer. Simul.
117 2023 106968 4512468 ( ), Paper No. . MR BSW S.
[23]
Bickel、B. Schleich 和 S. Wartzack，一种基于对象投影、预训练深度学习模型和自编码器的三维机械部件新型形状检索方法，Comput.-Aided Des. 154 (2023), Paper No. 103417. MR 4492099[CC21] Zhixiang Chen 和 Feilong Cao，具有简单架构和逼近能力的前馈神经网络构造，Math. Methods Appl. Sci. 44 (2021), no. 2, 1788–1795. MR 4185345[CCK22] Lucian Coroianu、Danilo Costarelli 和 Uğur Kadak，由S形激活函数的渐近行为所导出的神经网络算子的定量估计，Mediterr. J. Math. 19 (2022), no. 5, Paper No. 211, 25. MR 4476907[CKM22] Sitan Chen、Adam R. Klivans 和 Raghu Meka，深度ReLU网络学习是固定参数可处理的，2021 IEEE 第62届计算机科学基础年会——FOCS 2021，IEEE Computer Soc., Los Alamitos, CA, [2022] ©2022, pp. 696–707. MR 4399726[CL22] Hengjie Chen 和 Zhong Li，关于一个主函数在深度神经网络中应用的札记，Int. J. Wavelets Multiresolut. Inf. Process. 20 (2022), no. 4, Paper No. 2150058, 18. MR 4458444[CST01] Nello Cristianini 和 John Shawe-Taylor，《支持向量机及其他基于核的学习方法导论》，重印版，剑桥: 剑桥大学出版社，2001（英文）。[DWZ+23] Zeyu Dong、Xin Wang、Xian Zhang、Mengjie Hu 和 Thach Ngoc Dinh，离散时间高阶切换神经网络的全局指数同步及其在多通道音频加密中的应用，Nonlinear Anal. Hybrid Syst. 47 (2023), Paper No. 101291. MR 4491367[Eng96] Miroslav Engliš，复域上的Berezin量子化与再生核，Trans. Amer. Math. Soc.
348 1996 2, 411 479 1340173 ( ), no. – . MR
[23]
GK Philipp Grohs 和 Gitta Kutyniok（编），《深度学习的数学方面》，剑桥大学出版社，剑桥，2023。MR 4505882[GKNV22] Rémi Gribonval、Gitta Kutyniok、Morten Nielsen 和 Felix Voigtlaender，深度神经网络的逼近空间，Constr. Approx. 55 (2022), no. 1, 259–367. MR 4376564[GKP20] Ingo Gühring、Gitta Kutyniok 和 Philipp Petersen，在 $W^{s,p}$ 范数下深度ReLU神经网络逼近的误差界，Anal. Appl. (Singap.) 18 (2020), no. 5, 803–859. MR 4131039[GPR+21] Moritz Geist、Philipp Petersen、Mones Raslan、Reinhold Schneider 和 Gitta Kutyniok，利用深度神经网络数值求解参数化扩散方程，J. Sci. Comput. 88 (2021), no. 1, Paper No. 22, 37. MR 4268857[Han16] Muhammad Hanif，前馈人工神经网络的高斯-牛顿方法，Int. J. Math. Comput. 27 (2016), no. 3, 132–147. MR 3457582[HL23] Yuecai Han 和 Nan Li，一种用于多次停止的新型深度神经网络算法及其在期权定价中的应用，Commun. Nonlinear Sci. Numer. Simul. 117 (2023), Paper No. 106881. MR 4496811

[HST10] David R. Hardoon 和 John Shawe-Taylor，利用结构化标签分解用于神经科学数据的张量核支持向量机，Mach. Learn. 79 (2010), no. 1-2, 29–46. MR 3108145[HSTHD11] Zakria Hussain、John Shawe-Taylor、David R. Hardoon 和 Charanpal Dhanjal，正交匹配追踪算法的设计与泛化分析，IEEE Trans. Inform. Theory 57 (2011), no. 8, 5326–5341. MR 2849119[JR22] Arnulf Jentzen 和 Adrian Riekert，关于训练具有 ReLU 激活且目标函数为常数的人工神经网络时随机梯度下降收敛性的证明，Z. Angew. Math. Phys. 73 (2022), no. 5, Paper No. 188, 30. MR 4468133[JR23] ，训练具有ReLU激活的人工神经网络时梯度流的收敛性分析，J. Math. Anal. Appl. 517 (2023), no. 2, Paper No. 126601, 43. MR 4473797[JT18] Palle E.T. Jorgensen 和 Feng Tian，随机符号测度导出的高斯过程的对偶性，第2章，第23–56页，John Wiley & Sons, Ltd, 2018。[JT20] Palle E. T. Jorgensen 和 Feng Tian，谱对与正定温和分布，《分形上的分析、概率与数理物理》，Fractals Dyn. Math. Sci. Arts Theory Appl., vol. 5, World Sci. Publ., Hackensack, NJ, [2020] ©2020, pp. 223–241. MR 4472250[JT21] Palle Jorgensen 和 James Tian，无限维分析——希尔伯特空间中的算子；表示论视角下的随机演算与对偶理论，World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, [2021] ©2021。MR 4274591[JT22] ,再生核及其相关特征空间的选择，以 L^2-空间的形式呈现，J. Math. Anal. Appl. 505 (2022), no. 2, Paper No. 125535, 31. MR 4295177[KG22] Wei Kang 和 Qi Gong，前馈神经网络与复合函数及其在动力系统中的应用，SIAM J. Control Optim. 60 2, 786 2022 813 4395164 ( ), no. – . MR[KLLW21] Shi-Lei Kong、Ka-Sing Lau、Jun Jason Luo 和 Xiang-Yang Wang，迭代引出的双曲图及其在分形中的应用，《流形、分形与图上的分析和偏微分方程》，Adv. Anal. Geom., vol. 3, De Gruyter, Berlin, [2021] ©2021, pp. 143–181. MR 4320089[KLW20] Shi-Lei Kong、Ka-Sing Lau 和 Ting-Kam Leonard Wong，齐次型紧空间上的随机游走与诱导狄利克雷型，《分形上的分析、概率与数理物理》，Fractals Dyn. Math. Sci. Arts Theory Appl., vol. 5, World Sci. Publ., Hackensack, NJ, [2020] ©2020, pp. 273–296. MR 4472252[Kut20] Gitta Kutyniok，讨论：“使用具有ReLU激活函数的深度神经网络进行非参数回归” [*4134774*], **48** 2020 4, MR Ann. 1902 1905 4134776
Statist. ( ), no. – . MR[MCA20] Adrian Moldovan、Angel Caţaron 和 Răzvan Andonie，利用传递熵加速前馈神经网络的学习，Entropy 22 (2020), no. 1, Paper No. 102, 19. MR 4072078[MK23] Krzysztof Martyn 和 Mił osz Kadziński，用于多准则决策分析的深度偏好学习，European J. Oper. Res. 305 (2023), no. 2, 781–805. MR 4503771[MM23] Thomas Merkh 和 Guido Montúfar，随机前馈神经网络：通用逼近，《深度学习的数学方面》，剑桥大学出版社，剑桥, 2023, 267 314 4505888 pp. – . MR OSZ Joost A. A. [22]
Opschoor、Christoph Schwab 和 Jakob Zech，高维深度学习：用于贝叶斯偏微分方程反演的ReLU神经网络表达，《偏微分方程的优化与控制——不确定性量化、开环与闭环控制及形状优化》，Radon Ser. Comput. Appl. Math., vol. 29, De Gruyter, Berlin, [2022]©2022, pp. 419–462. MR 4409717[PORSTS21] María Pérez-Ortiz、Omar Rivasplata、John Shawe-Taylor 和 Csaba Szepesvári，神经网络更紧致的风险证书，J. Mach. Learn. Res. 22 (2021), Paper No. 227, 40. MR 4329806

[RSSST06] Juho Rousu、Craig Saunders、Sandor Szedmak 和 John Shawe-Taylor，层次多标签分类模型的核方法，《机器学习研究杂志》**7**（2006），1601–1626。MR 2274418[SHO22] Alexandre Smirnov、Boumediene Hamzi 和 Houman Owhadi，深度学习中训练权重的平均场极限：动力系统视角，《多洛米蒂近似研究札记》**15** （2022），专刊：为庆祝 Robert Schaback 75 岁生日而作，125–145。MR 4500409[STWCK05] John Shawe-Taylor、Christopher K. I. Williams、Nello Cristianini 和 Jaz Kandola，关于核*PCA*的*Gram*矩阵特征谱与泛化误差，《IEEE信息论汇刊》 **51** 2005 7, 2510 2522 2246374 （），[83]
第 – 期。MR Sza F. H. Szafraniec，正定核的插值与支配，《具有再生核的希尔伯特空间》，雅盖隆大学出版社，克拉科夫，2004年（波兰语）。，《数学讲义》，第1014卷，施普林格，柏林，1983年，第291–295页。MR 738131[Sza04] ,《具有再生核的希尔伯特空间》（波兰语），雅盖隆大学出版社，克拉科夫，2004年。[Sza15a] Franciszek Hugon Szafraniec，再生核性质及其空间：或多或少标准的应用示例，《算子理论》（含51幅图和2张表，共2卷），巴塞尔：施普林格，2015年，第31–58页（英文）。[Sza15b] , 再生核性质及其空间：基础，《算子理论》（含51幅图和2张表，共2卷），巴塞尔：施普林格，2015年，第3–30页（英文）。[Sza21] ,重振*Pedrick*的方法以研究再生核希尔伯特空间，《复分析与算子理论》**15** （2021），第4期，论文编号66，共12页。MR 4250453[ZC23] Guodong Zhang 和 Jinde Cao，时滞模糊惯性不连续神经网络固定时间/预设时间同步的新结果：非降阶方法，《应用数学与计算》**440**（2023），论文编号127671。MR 4505410

(Palle E.T. Jorgensen) 数学系，爱荷华大学，爱荷华州 City，爱荷华州 52242-1419 ，美国
电子邮箱：palle-jorgensen@uiowa.edu

(Myung-Sin Song) 数学与统计学系，南伊利诺伊大学爱德华兹维尔分校，爱德华兹维尔，伊利诺伊州 62026，美国电子邮箱：msong@siue.edu

(James F. Tian) 《数学评论》，安娜堡第四街416号，密歇根州 48103-4816，美国
电子邮件地址：jft@ams.org