arXiv:2603.02973v1 [cs.LG] 3 Mar 2026

论神经网络超水平集的拓扑

BAHMAN GHARESIFARD

摘要。我们证明，激活函数满足 Riccati 型常微分方程条件的神经网络——该假设源于近期在一致拓扑中的通用逼近结果——在解析域上产生 Pfaffian 输出，其格式仅由架构控制。因此，超水平集以及神经网络参数化向量场的李括号秩下降轨迹，在拓扑复杂度上（特别是总 Betti 数上）具有仅依赖于架构的界，且该界对所有权重一致成立。

1. 引言

神经网络通常部署于如下情形：显著的几何对象并非原始标量得分 $F : \mathbb{R}^d \to \mathbb{R}$ 本身，而是一个 阈值区域 （或 超水平集）

$$\mathcal{S}_{\geq\tau}(F) := \{x \in V : F(x) \geq \tau\},$$

对于固定阈值 $\tau \in \mathbb{R}$ 和域 $V \subset \mathbb{R}^d$。此类集合编码了关于输入—输出映射的全局几何信息，并且先验地可以展现丰富的拓扑，例如许多连通分支和高维孔洞，即使当 F 光滑时亦然。

大量文献通过决策复杂度的几何和组合替代指标来研究神经网络的容量。一条经典研究路线通过计算线性区域的数量及其对深度和宽度的依赖来量化表达能力 [11, 13, 12], ；与之互补的研究则强调深度驱动的振荡和深度分离现象 [17]。另一种观点，尤其由 [2, 7, 3], 倡导，是通过决策区域的拓扑不变量来衡量分类器复杂度，并通过上下界比较浅层和深层架构。我们的工作与这一拓扑视角密切相关，它探讨的是$\mathcal{S}_{\geq\tau}(F)$是否会变得任意复杂。

然而，与现有文献中的结果不同，我们从通用逼近理论出发提出一种结构性视角，解释了为何一旦架构固定，这些拓扑界就能对所有权重统一成立。更具体地说，我们证明，对于一大类由*Riccati* 型微分约束刻画的光滑激活函数，在任意固定的解析域上，网络输出都属于一个温和的 Pfaffian 类，其格式 仅由架构参数（深度和宽度）以及 Riccati 指数控制。与 [2], 一致，这些结果给出了在固定解析域上，文档后文定义的 $\mathcal{S}_{\geq\tau}(F)$的总 Betti 数的显式架构专属界。

Riccati 假设发挥着核心作用。它要求对于某个 $r \in \mathbb{N}_0$，激活函数的 r 阶导数在其解析域上满足一个 Riccati 型二次微分方程。这一条件并非凭空设定：在 [14, 15, 16], 中，研究证明，激活函数本身或其某个有限阶导数满足这种二次微分方程性质，是深度残差／流模型在一致拓扑下实现通用逼近的一项充分结构性假设。事实证明，许多常用的光滑激活函数都满足这一性质。

第二项贡献在于，同一结构机制可以从标量决策区域推广到神经网络参数化向量场时产生的几何轨迹。这种情况例如出现在：学习或控制更新律通过改变网络权重生成一族动力学，从而每个

作者的研究工作得到加拿大自然科学与工程研究理事会的资助。

参数选择都会在固定状态空间上得到一组不同的向量场。尽管我们将在后文详细介绍相关构造，但这里先简要明确这一设定，以说明几点：固定m实解析向量场 $X_1, \dots, X_m$于 $V \subset \mathbb{R}^d$上，其坐标分量由具有固定架构的神经网络实现，并令 $\Delta^k(z)$表示由 $\{X_1, \dots, X_m\}$的所有长度为k的迭代李括号张成的空间。我们考虑秩下降分层

$$\mathcal{Z}_{k,\rho} = \{z \in V : \dim \Delta^k(z) \le \rho\}.$$

这类集合刻画了括号生成方向无法达到指定维数阈值的区域，在几何控制 [9] 和次黎曼几何 [1]中十分常见。特别是，它们在刻画神经网络动力学的输出可达集方面发挥着核心作用。我们证明，在相同的 Riccati激活假设下，这些分层的拓扑也具有明确的、与权重无关的界；这些界仅取决于（d， m， k， ρ）和网络架构。据作者所知，现有研究中尚未出现针对神经网络参数化向量场所产生的李括号秩下降轨迹的贝蒂数权重一致界。

从技术层面来说，证明的要点是表明，在任意解析区域上，网络输出（以及在控制场景中，向量场分量及由其构造的李括号表达式）都属于 Pfaffian 函数类，其格式 仅取决于网络架构和 Riccati 指数。随后，类似于 [2], 中所用的经典 Pfaffian复杂性定理便可推出所需的零点数和贝蒂数一致界 [10, 18, 6]。

2. 神经网络超水平集：一个拓扑问题

考虑一个标准的前馈神经网络，其输入维度为 $n_0 := d \in \mathbb{N}$，其中 $\mathbb{N}$ 是正整数集，深度为 $L \in \mathbb{N}$，各层宽度为 n_1、.....、$n_L \in \mathbb{N}$。对于每个 $\ell \in \{1, \dots, L\}$，令

$$W^{(\ell)} \in \mathbb{R}^{n_\ell \times n_{\ell-1}}, \qquad b^{(\ell)} \in \mathbb{R}^{n_\ell}$$

表示第 ℓ 层的权重矩阵和偏置向量。令 $\sigma : \mathbb{R} \to \mathbb{R}$ 为一个激活函数，并定义其逐分量扩展 $\Sigma_\ell : \mathbb{R}^{n_\ell} \to \mathbb{R}^{n_\ell}$ 如下:

$$\Sigma_\ell(z) := \big(\sigma(z_1), \dots, \sigma(z_{n_\ell})\big).$$

给定输入 $x \in \mathbb{R}^d$，令 $h^{(0)}(x) := x \in \mathbb{R}^{n_0}$，并对每个 $\ell \in \{1, \dots, L\}$递归定义:

$$h^{(\ell)}(x) = \Sigma_\ell\big(W^{(\ell)} h^{(\ell-1)}(x) + b^{(\ell)}\big) \in \mathbb{R}^{n_\ell}. \tag{2.1}$$

为简单起见，我们考虑如下形式的标量输出映射 $F : \mathbb{R}^d \to \mathbb{R}$：

$$F(x) := c_0 + c^\top h^{(L)}(x), \tag{2.2}$$

其中 $c_0 \in \mathbb{R}$ 且 $c \in \mathbb{R}^{n_L}$。下文的结构性部分可逐分量扩展至向量值输出，以下陈述适用于标量坐标或其标量线性组合。

现在我们聚焦于本文所考虑的激活函数类别。在 [15], 中，以及在由神经常微分方程建模的深度神经网络的通用逼近背景下，以下激活函数类别已被证明足以实现一致拓扑下的通用逼近。接下来我们回顾这一激活函数类别。

定义 2.1. 对于 $r \in \mathbb{N}_0$，我们称 $\sigma : \mathbb{R} \to \mathbb{R}$ 属于 $\mathcal{A}_{\text{quad},\, r}$，如果它是非递减的，并且存在常数 a_0、a_1、$a_2 \in \mathbb{R}$，满足 $a_2 \neq 0$，使得定义 ***Riccati***常微分方程 (Riccati 常微分方程):

$$\zeta'(t) = a_0 + a_1\zeta(t) + a_2\zeta(t)^2, \tag{2.3}$$

其中 t 位于解有定义的开集中，我们有

$$\zeta(t) = \frac{d^r \sigma}{dt^r}(t).$$

我们将 r 称为$\sigma \in \mathcal{A}_{\text{quad},\, r}$的Riccati 指数。

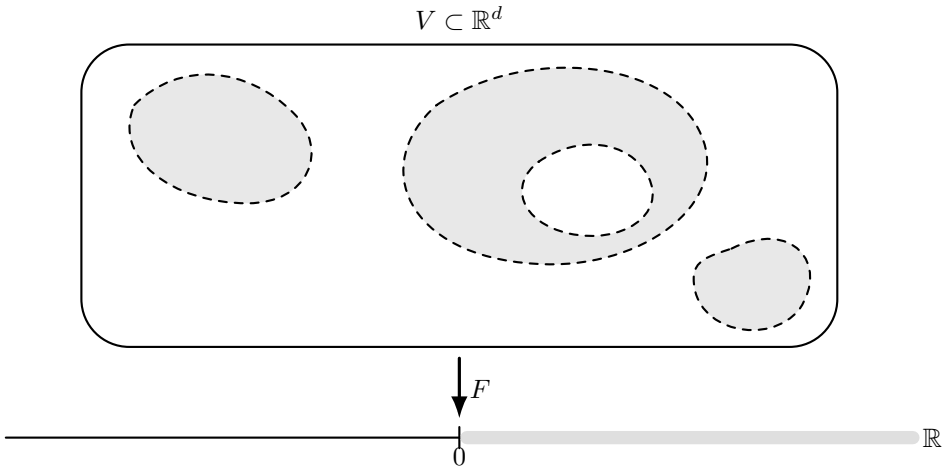


图 1. 示意图描绘了标量映射 $F\colon V \to \mathbb{R}$ 及其决策区域 $D = \{F \geq 0\} = F^{-1}([0,\infty))$。虚线表示零水平集。

换言之，要么 σ ，要么其 r阶导数满足 Riccati 常微分方程 (2.3)。有趣的是，文献中使用的一大类激活函数都满足这一条件，包括逻辑函数、双曲正切函数和 Softplus 函数。此外，其他常用函数（如 ReLU 和 GeLU）也可以在这一类中得到很好的逼近。

显然，上述定义要求 $\sigma \in C^{r+1}$。注意，实解析常微分方程（ODE）的解，即右端为实解析函数的 ODE 的解，在其作为经典解存在的任何开区间上都是实解析的；实解析函数的反导函数也是实解析的。因此，在满足式 (2.3) 且 ζ有定义（即不存在爆破）的任何开区间 I 上，函数 ζ，$\sigma^{(j-1)}$，……，σ在 I 上都是实解析的。除非另有说明，以下所有结果均局部地 在所有相关表达式均为解析函数的任意开域上陈述。

现在，我们定义本文感兴趣的主要目标之一。为便于阅读，更复杂的参数化向量场情形将留到第 3.1 小节讨论。

定义 n 2.2. 对于定义在区域 $V \subseteq \mathbb{R}^d$ 上的实值函数 F，超水平集定义为 s

$$\mathcal{S}_{\geq 0}(F) := \{x \in V :\ F(x) \geq 0\}.$$

现在，我们先非正式地陈述本文关注的问题之一，将更复杂的参数化向量场讨论留待后文：

问题陈述： 我们关注的是神经网络标量输出函数 (2.2) 的超水平集的拓扑结构。例如，对于二元分类问题，其中实值评分 F 的符号 用于生成取值于 $\{+1,-1\}$ 的标签，我们想知道集合 $\mathcal{S}_{\geq 0}(F)$ 是否可能具有任意复杂的结构，并包含许多连通分支。

3. Main Results

我们先介绍一种特殊情形，其中 $d = 1$；这样做是因为该结果更容易解释，同时也为后续结果奠定基调。设 $I \subset \mathbb{R}$ 为一个开区间。给定一个不恒为零的实解析函数 $F\colon I \to \mathbb{R}$，我们记

$$\mathrm{Zeros}(F; I) := \big|\{x \in I :\ F(x) = 0\}\big|,$$

I中的不同实零点的数量。

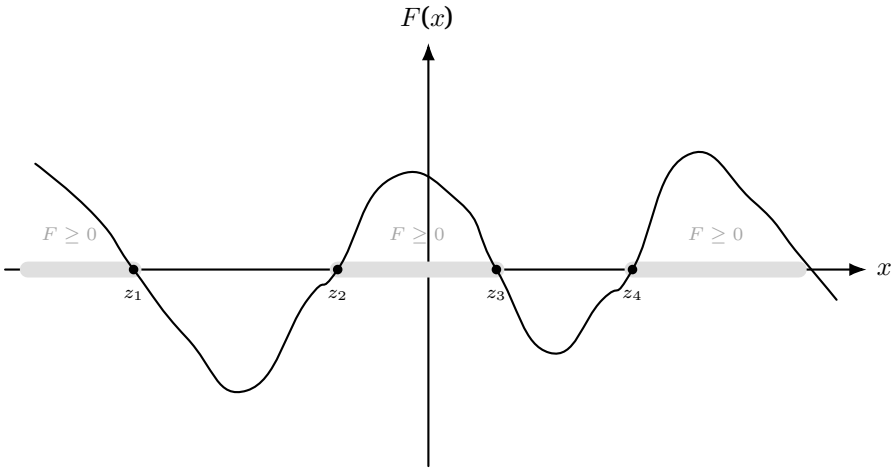


图2. 假设具有解析性，每次符号变化只能发生在 F 的零点处，因此区间分支的数量受 Zeros(F; I) 控制。

命题3.1。 设 $d = 1$ ，且 $F:\ I \to \mathbb{R}$ 是如 (2.1) 中所述、宽度为 n_1、...、 n_L 和 $\sigma \in \mathcal{A}_{\text{quad},\ r}$ 的深度-L 神经网络在开区间 I 上的输出，其中 F 是解析的。令

$$R := (r+2)\sum_{\ell=1}^{L} n_\ell. \tag{3.1}$$

那么，要么 $F \equiv 0$ 在 I 上，或者

$$\text{Zeros}(F; I) \ \leq\ 2^{\frac{R(R+1)}{2}} C_I (1+L)^{R+1},$$

其中， $C_I > 0$ 是一个取决于区间 I、但与网络权重和偏置无关的常数。特别地，超水平集 $\mathcal{S}_{\geq 0}(F)$ 是至多由 $2^{R}{}_{(R+1)^2}\ \ C_I(1+L)^{R+1}$ 个区间组成的并集，这一结论对所有权重值均成立。

这里有几点需要说明。在一维情形中，分类器的表达能力 自然可用沿输入直线发生的标签翻转次数来刻画，也就是 F 的符号变化次数。已有多项研究（例如 [13]）通过考察网络对一维输入路径的扭曲程度来量化表达能力，例如计算诱导表示轨迹的弧长。与此不同，上述一维结果直接控制了沿输入直线的一种拓扑 复杂度：$x \mapsto \text{sign}(F(x))$ 的标签翻转次数，等价地说，即 F 的零点数，从而也就是 $\{F \geq 0\}$ 的区间分支数。这种互补的拓扑视角已在文献中得到倡导；例如参见 [7, 3], ，并应将其视为一种最坏情况 下的表达能力界；对于给定架构，它为所有参数选择下可能出现的最大振荡提供了上界。

上述界对权重一致成立：对于固定的架构和 $\sigma \in \mathcal{A}_{\text{quad},\ r}$，无论如何选择参数，都无法在 I 上产生任意多次符号变化。这种一致性也解释了为何该界预期会较大：它必须同时适用于每一种参数选择，而不是描述训练后网络的典型行为。如引言所述，该结果表明，在 [2] 中观察到的拓扑控制类型可推广至所有满足某个 Riccati 指数 r所对应的 Riccati型常微分方程 (2.3) 的激活函数。正如 [2], 中所指出的，增加深度 L可以显著提升最坏情况下的表达能力，因为这会增大指数项的底数。

尽管这或许显而易见，我们仍要强调，这种复杂度概念不同于 VC维数等统计容量度量：这里我们控制的是决策区域的全局几何特征，而不是有限样本的打散。最后，若满足以下条件，则可确保 F 在 I 上实解析：对于

每个层 ℓ 和神经元 k，我们都要求相关量的取值位于一个开区间内，且 σ 在该区间上实解析；也就是说，这些取值远离 Riccati 解的任何奇点或爆破点（如果存在）。全文中我们都默认将范围限制在这样的区间内。

当然，连通分支的数量并不能反映超水平集的所有拓扑特征。因此，我们使用贝蒂数（见定义 4.3），它们可量化集合在不同维度上的拓扑，例如测量分支的数量以及是否存在高维孔洞。

现在我们可以陈述命题 3.1 的一个推广。

定理 3.2. 假设 $d \geq 1$ 且 $F\colon\ V \to \mathbb{R}$ 是如 (2.1)-(2.2) 所示的深度为 L 的神经网络的输出，其宽度为 $n_1, \ldots, n_L$ 且 $\sigma \in \mathcal{A}_{\mathrm{quad},r}$。假设 $V \subset \mathbb{R}^d$ 是一个开域，在该域上 F 是实解析的，并令 R 如 (3.1) 所示。那么

$$\mathrm{Betti}\big(\mathcal{S}_{\geq 0}(F)\big) \ \leq\ \mathfrak{B}_V(d,R,L),$$

其中

$$\mathfrak{B}_V(d,R,L) := 2^{\frac{R(R-1)}{2}}\ C_V\ \Big(d + \min\{d,R\}(1+2L)\Big)^{d+R}.$$

这里 $C_V > 0$ 是一个取决于开域 V、但与网络权重和偏置无关的常数。特别地，决策区域的连通分支数量至多为 $\mathfrak{B}_V(d,R,L)$。

3.1. 控制几何。 在许多值得关注的场景中，神经网络用于对状态空间上的向量场（或受控动力学）进行参数化。具体而言，固定一种架构并改变权重，就会得到一族向量场。例如，可以有 m 个向量场 $X_1,\ldots,X_m$ 定义在 $V \subset \mathbb{R}^d$ 上，其坐标函数 $X_{i,p}$ 由神经网络实现；例如，研究通用逼近性质 [14]时，这一点就十分重要。与此类向量场族相关的一个基本几何对象，是这些向量场及其迭代李括号共同张成的空间：在每个点 z，括号张成空间 $\Delta^k(z)$刻画了通过组合向量场能够生成的方向，组合所用的括号长度不超过 k。 $\Delta^k(z)$的维数是衡量系统在 z附近能够产生多少个方向的局部指标，而秩下降轨迹$\mathcal{Z}_{k,\rho}$ 则记录了该张成空间未达到指定维数阈值的位置。

与超水平集视角的关系。 在前文研究的分类场景中，人们通过由单个标量评分函数F定义的阈值区域的拓扑，来量化最坏情况下的分类器复杂度。在此处考虑的控制场景中，对应的几何对象是秩下降轨迹

$$\mathcal{Z}_{k,\rho} = \{z \in V :\ \dim \Delta^k(z) \leq \rho\}, \tag{3.2}$$

其中， $\Delta^k(z)$ 表示由向量场 X_1、……、 X_m 的所有长度不超过 k 的李括号张成的空间。从这个意义上说，所采用的观点保持不变，但所测量的量发生了变化：我们测量的是在给定括号深度 k 下，无法生成足够多线性无关方向的程度。

设 $V \subset \mathbb{R}^d$ 为开集，并设 X_1、……、 X_m 为 V 上的实解析向量场，

$$X_i(z) = \sum_{p=1}^{d} X_{i,p}(z)\,\partial_{z_p}.$$

固定 $k \in \mathbb{N}$，并令 $\mathcal{B}_k$ 为由 $\{X_1,\ldots,X_m$ 构造的所有长度不超过 k 的迭代李括号组成的有限集 $\}$。定义

$$\Delta^k(z) := \mathrm{span}\{Y(z) :\ Y \in \mathcal{B}_k\} \subset \mathbb{R}^d,$$

并令 $A_k(z)$ 为其列向量是 $Y(z)$ 的坐标向量的 $d \times |\mathcal{B}_k|$ 矩阵。对于 $\rho \in \{0,1,\ldots,d\}$，式 (3.2) 中的秩下降轨迹可写为

$$\mathcal{Z}_{k,\rho} = \{z \in V :\ \mathrm{rank}\, A_k(z) \leq \rho\}.$$

设 M_1，……，$M_{s_{k,\rho}}$ 表示 A_k 的全部 $(\rho+1) \times (\rho+1)$ 子式所构成的列表，其中

$$s_{k,\rho} := \binom{d}{\rho+1}\binom{|\mathcal{B}_k|}{\rho+1}.$$

定理 3.3. 假设每个系数函数 $X_{i,\ p}:\ V \to \mathbb{R}$ 在 V 上由具有固定架构 $(d,\ L,\ n_1,\ \ldots\ldots,\ n_L)$ 及激活函数 $\sigma \in \mathcal{A}_{\mathrm{quad},\ r}$ 的前馈神经网络实现，并且所有这些系数函数在 V 上均为实解析的。令

$$s_{k,\rho} := \binom{d}{\rho+1}\binom{|\mathcal{B}_k|}{\rho+1}.$$

Then

$$\mathrm{Betti}(\mathcal{Z}_{k,\rho}) \ \le\ \mathfrak{R}_V\Big(d, m, k, \rho;\, L, n_1, \ldots, n_L, r\Big),$$

其中

$$\mathfrak{R}_V\Big(d, m, k, \rho;\, L, n_1, \ldots, n_L, r\Big) := 2^{\frac{R_k(R_k-1)}{2}}\ (s_{k,\rho})^d\ C_V\ \Big(d\beta_k + \min\{d, R_k\}\alpha_k\Big)^{d+R_k},$$

其中 $C_V > 0$ 是一个仅取决于固定区域 V、与所有网络权重和偏置无关的常数； R_k、α_k、β_k 是对所有子式 M_i、$i \in \{1, \ldots, s_{k,\rho}\}$均一致成立的常数。

定理 3.3 给出了与权重无关的控制，刻画秩下降分层的几何复杂度$\mathcal{Z}_{k,\ \rho}$。注意，由于 $\mathcal{B}_k \subseteq \mathcal{B}_{k+1}$，对所有 $z \in V$，都有 $\Delta^k(z) \subseteq \Delta^{k+1}(z)$，因此

$$\mathcal{Z}_{k+1,\rho} \subseteq \mathcal{Z}_{k,\rho}, \qquad \text{for all } k, \rho, \tag{3.3}$$

因此，增大 k 只会缩小退化集。参数 k 应理解为用于衡量秩生成的选定括号深度；然而，不同的权重选择可能在不同深度达到满秩。

4. Proof of main results

W我们在介绍一些数学预备知识后，证明本文的主要结果 s

4.1. 数学预备知识。 我们主要证明的关键思路源于这样一个事实：在我们对激活函数类所作的假设下，神经网络输出函数的行为是温和的，因为它们属于“Pfaffian 函数”这一类，我们很快就会回顾其定义，见 [10, 6, 18]。随后，我们利用关于由这类函数定义的超水平集的拓扑性质的一些经典结果，从而得出结论。我们先定义这一类函数。

定义 4.1。 设 $V \subset \mathbb{R}^d$ 为一个开子集。若对于每个 $i \in \{1, \ldots, R\}$ 和每个坐标指标$p \in \{1, \ldots, d\}$，都存在一个总次数至多为 α 的多项式 $P_{i,p}(x, y_1, \ldots, y_i)$，使得对所有 $x \in V$，则称 $f_i\colon V \to \mathbb{R}$（其中每个 f_i 在 V上实解析）构成一个次数为 α 的 Pfaffian 链 α，

$$\frac{\partial f_i}{\partial x_p}(x) = P_{i,p}\big(x, f_1(x), \ldots, f_i(x)\big).$$

实解析函数 $F\colon\ V \to \mathbb{R}$ 是格式为 (d, R, α, β) 的 ***Pfaffian*** 函数，当且仅当

$$F(x) = Q\big(x, f_1(x), \ldots, f_R(x)\big),$$

存在某个 ***Pfaffian*** 链$(f_1, \ldots, f_R)$，其次数为 α，以及某个次数至多为 β 的多项式 Q。

我们举例说明。设 $V = \mathbb{R}$，并定义 $f_1(x) = e^x$。那么

$$\frac{df_1}{dx}(x) = e^x = f_1(x)$$

此后，(f_1) 是一个次数为 $\alpha = 1$ 的 Pfaffian 链。因此，任意多项式

$$F(x) = Q\big(x, f_1(x)\big) = Q(x, e^x)$$

在 $\mathbb{R}$ 上是 Pfaffian 函数；例如， $F(x) = x^2 + e^x$ 是格式为 (1，1，1，2) 的 Pfaffian 函数。

欧拉伽马函数是一个实解析函数的例子，但它在正实数的任何开子区间上都不是 Pfaffian 函数，参见 [4]。

我们关注由一组 Pfaffian 函数定义的子集，下面将对此进行概述。

定义 4.2。 令 $V \subset \mathbb{R}^d$ 为开集。若存在 ***Pfaffian*** 函数G_1、.....、G_s： $V \to \mathbb{R}$，使得 S 可以由形如以下形式的集合得到，则称集合 $S \subset V$ 在 V上是半 Pfaffian的

$$\{x \in V : \ G_i(x) = 0\}, \qquad \{x \in V : \ G_i(x) \geq 0\}, \quad \text{and} \quad \{x \in V : \ G_i(x) \leq 0\},$$

其中 $i \in \{1, \ldots, s\}$，通过有限次取并集和交集得到。

这里指出，文献中半 Pfaffian 的定义比此处给出的更一般，而此处的定义已足以满足我们的目的 [18, 6], 。一个与本研究相关的例子是，当 G: $V \to \mathbb{R}$ 是 Pfaffian 函数，且 $V \subset \mathbb{R}^d$ 是一个开集时，超水平集 $\{x \in V : \ G(x) \geq 0\}$ 是 V 上的半 Pfaffian 集。

最后，我们回顾拓扑空间的贝蒂数的定义。完整介绍请参阅 [8, 第 2]章。

定义 4.3. 设 $S \subset \mathbb{R}^d$ 是一个拓扑空间。对于每个 $i \geq 0$， i阶贝蒂数（取 $\mathbb{Z}_2$系数）为

$$b_i(S) := \dim_{\mathbb{Z}_2} H_i(S; \mathbb{Z}_2),$$

其中 $H_i(S; \mathbb{Z}_2)$ 是 i阶奇异同调群。总贝蒂数 为

$$\text{Betti}(S) := \sum_{i=0}^{d} b_i(S).$$

我们回顾定理 3.4] 中的下一个重要经典结果。

定理 4.4. 设 $V \subset \mathbb{R}^d$ 是一个开域， $S \subset V$ 是一个半 ***Pfaffian*** 集，由定义在 V 上、格式为 (d, R, α, β) 的 s ***Pfaffian*** 函数描述。那么

$$\text{Betti}(S) \ \leq \ 2^{\frac{R(R-1)}{2}} \ s^d \ C_V \ \Big(d\beta + \min\{d, R\}\alpha\Big)^{d+R},$$

其中， $C_V > 0$ 是一个常数，仅取决于固定域 V 和 ***Pfaffian*** 链，而与定义 S 的 ***Pfaffian*** 函数中的系数/参数无关。

4.2. **主要证明**。 我们首先证明命题3.1。根据上一小节中确立的预备结果，只需证明在关于激活函数的假设下，给定神经网络的输出函数是 Pfaffian 函数。

命题 4.5. 设 $\sigma \in \mathcal{A}_{\text{quad},r}$ 的 ***Riccati*** 指数为 r。考虑一个深度为 L 的神经网络 (2.1)–(2.2) ，其宽度为 n_1、.....、n_L，输出为标量 F。设 $V \subset \mathbb{R}^d$ 是一个开集，在该开集上所有仿射层输入

$$s^{(\ell)}(x) := W^{(\ell)} h^{(\ell-1)}(x) + b^{(\ell)} \in \mathbb{R}^{n_\ell}, \qquad \ell = 1, \ldots, L,$$

取值于一个开区间，在该区间上 σ 是实解析的。于是， F 在 V 上是 ***Pfaffian*** 的，其格式为

$$(d, R, \alpha, \beta) = \Big(d, \ (r+2)\sum_{\ell=1}^{L} n_\ell, \ 1 + 2L, \ 1\Big). \tag{4.1}$$

证明。 固定 $\ell \in \{1, \ldots, L\}$，并将 $s^{(\ell)}(x) \in \mathbb{R}^{n_\ell}$ 和 $h^{(\ell)}(x) \in \mathbb{R}^{n_\ell}$ 的坐标分别记为 $s_k^{(\ell)}(x)$ 和 $h_k^{(\ell)}(x)$。对于每个神经元 (ℓ, k) 以及每个整数 $q \in \{0, 1, \ldots, r\}$，我们定义辅助函数

$$u_{k,q}^{(\ell)}(x) \ := \ \frac{d^{\,q}\sigma}{dt^{\,q}}\Big|_{t = s_k^{(\ell)}(x)},$$

对于 $q = 0$，这意味着

$$u_{k,0}^{(\ell)}(x) = \sigma\big(s_k^{(\ell)}(x)\big) = h_k^{(\ell)}(x),$$

其中 r 是 Riccati 指数。

为验证 F 是 Pfaffian，我们将通过明确构造一个 Pfaffian 链来进行说明：将所有仿射层输入 $s_k^{(\ell)}(x) = (W^{(\ell)} h^{(\ell-1)}(x) + b^{(\ell)})_k$ 列为定义在V上的标量值函数，并将对这些标量求取 F 导数所得的函数一并列出。

首先，对于每一层 $\ell \in \{1,\dots,L\}$ 和每个神经元索引 $k \in \{1,\dots,n_\ell\}$，我们将相关函数汇集到以下块中

$$\mathcal{F}_{\ell,k} := \big(s_k^{(\ell)},\ u_{k,r}^{(\ell)},\ u_{k,r-1}^{(\ell)},\dots,u_{k,0}^{(\ell)}\big), \tag{4.2}$$

其长度为 $r+2$。注意，每个块 $\mathcal{F}_{\ell,k}$ 都是与神经元 (ℓ,k) 相关联的实解析函数有限元组。我们将这些块中出现的标量函数排序并组成一个列表，使得对于每个列表元素 f_i 和每个坐标指标 p，导数 $\partial_p f_i$ 都可以写成 $(x, f_1, \dots, f_i)$ 的多项式，如 Pfaffian 链定义所要求的那样。

现在，我们按照上述方式，通过按以下字典序拼接这些块，构造有序列表 $(f_1,\dots,f_R)$：先按神经元指标递增顺序列出第 1 层的所有块，然后列出第 2 层的所有块，依此类推，直到第 L 层，即

$$\mathcal{F}_{1,1},\dots,\mathcal{F}_{1,n_1},\ \mathcal{F}_{2,1},\dots,\mathcal{F}_{2,n_2},\ \dots,\ \mathcal{F}_{L,1},\dots,\mathcal{F}_{L,n_L}.$$

在每个固定块 $\mathcal{F}_{\ell,k}$ 内部，我们保留内部顺序 $s_k \prec u_{k,r} \prec u_{k,r-1} \prec \cdots \prec u_{k,0}^{(\ell)}$。按照这一约定，属于第 ℓ 层的每个函数都排在来自第 $<\ell$ 层的所有函数之后；对于固定的 (ℓ,k)，函数 $u_k^{(\ell)}{}_{,q+1}$ 排在函数 $u_k^{(\ell)}{}_{,q}$ 之前。

由于每个块的长度为 $r+2$，我们有 $R = (r+2)\sum_{\ell=1}^{L} n_\ell$. 现在固定一个坐标指标 $p \in \{1,\dots,d\}$，并记 $\partial_p := \frac{\partial}{\partial x_p}$。由网络递归 (2.1)，我们有

$$\partial_p s_k^{(\ell)}(x) = \sum_{m=1}^{n_{\ell-1}} W_{km}^{(\ell)}\, \partial_p h_m^{(\ell-1)}(x) = \sum_{m=1}^{n_{\ell-1}} W_{km}^{(\ell)}\, \partial_p u_{m,0}^{(\ell-1)}(x), \tag{4.3}$$

对于 $\ell = 1$，我们有 $s_k^{(1)}(x) = \sum_{m=1}^{d} W_{km}^{(1)} x_m + b_k^{(1)}$，因此 $\partial_p s_k^{(1)}(x) = W_{kp}^{(1)}$。

此外，对于 $q \in \{0,\dots,r-1\}$，

$$u_{k,q}^{(\ell)}(x) = \frac{d^{\,q}\sigma}{dt^{\,q}}\Big|_{t=s_k^{(\ell)}(x)}.$$

现在使用链式法则，得到

$$\partial_p u_{k,q}^{(\ell)}(x) = \frac{d^{\,q+1}\sigma}{dt^{\,q+1}}\Big|_{t=s_k^{(\ell)}(x)}\, \partial_p s_k^{(\ell)}(x) = u_{k,q+1}^{(\ell)}(x)\, \partial_p s_k^{(\ell)}(x). \tag{4.4}$$

对于 $q = r$，令 $\zeta(t) := \frac{d^{\,r}\sigma}{dt^{\,r}}(t)$，于是 $u_{k,r}^{(\ell)}(x) = \zeta(s_k^{(\ell)}(x))$。根据定义 2.1，$\zeta$ 满足 (2.3)，即 $\zeta'(t) = a_0 + a_1\zeta(t) + a_2\zeta(t)^2$, 因此，再次应用链式法则，

$$\partial_p u_{k,r}^{(\ell)}(x) = \zeta'\big(s_k^{(\ell)}(x)\big)\, \partial_p s_k^{(\ell)}(x) = \big(a_0 + a_1 u_{k,r}^{(\ell)}(x) + a_2 (u_{k,r}^{(\ell)}(x))^2\big)\, \partial_p s_k^{(\ell)}(x). \tag{4.5}$$

现在我们提出下一个论断，由此即可得出结论。

子引理1。对于每个 $\ell \in \{1,\dots,L\}$，令 $D_\ell := 1 + 2\ell$。那么对于每个 $p \in \{1,\dots,d\}$ 以及每个列表元素 f_i，若其块 $\mathcal{F}_{\ell',k'}$ 在 (4.2) 中属于前 ℓ 层之一，即 $\ell' \le \ell$，则偏导数 $\partial_p f_i$ 可以表示为关于 $(x, f_1, \dots, f_i)$ 的一个多项式，其总次数至多为 D_ℓ。特别地，令 $\alpha := D_L = 1 + 2L$，则有序列表 $(f_1, \dots, f_R)$ 是 V 上的一个次数为 α 的 ***Pfaffian*** 链。

Pro子引理的证明。我们对 ℓ 使用归纳法证明该结论。先考虑基础情形 .

情形 $\ell = 1$。我们有 $\partial_p s_k^{(1)}(x) = W_{kp}^{(1)}$，这是一个常数（次数为0）。由(4.4)，对于 $q < r$，我们有 t

$$\partial_p u_{k,q}^{(1)}(x) = u_{k,q+1}^{(1)}(x)\, W_{kp}^{(1)},$$

它是关于 $(x, f_1, \dots, f_i)$ 的一次多项式。现在利用 (4.5) 可得

$$\partial_p u_{k,r}^{(1)}(x) = \big(a_0 + a_1 u_{k,r}^{(1)}(x) + a_2 (u_{k,r}^{(1)}(x))^2\big)\, W_{kp}^{(1)},$$

它是关于 $(x,\ f_1,\ \dots\dots,\ f_i)$ 的二次多项式。因此，第一层中列表元素的所有导数都是关于 $(x,\ f_1,\ \dots\dots,\ f_i)$ 的次数至多为 $D_1 = 3$ 的多项式。这证明了基础情形。

归纳步骤。假设该命题对 $\ell-1\geq 1$ 成立，即第一$\ell-1$层中每个列表元素 f_j 的 $\partial_p f_j$ 都是关于 $(x,\ f_1,\ \ldots\ldots,\ f_j)$ 的次数至多为 $D_{\ell-1}$ 的多项式。

固定第 ℓ 层中的一个神经元 $(\ell,\ k)$。根据归纳假设，每个 $\partial_p u_{m\ 0}^{(\ell-1)}\ ,\ (x)$ 都是关于较早列表元素、次数至多为 $D_{\ell-1}$ 的多项式，因此 (4.3) 将 $\partial_p s_k^{(\ell)}\ (x)$ 表示为这类多项式的线性组合。因此， $\partial_p s_k^{(\ell)}\ (x)$ 本身是关于较早列表元素、次数至多为 $D_{\ell-1}$ 的多项式。

现在，固定 $q\in\{0,\ldots,r-1\}$。由 (4.4)，我们有

$$\partial_p u_{k,q}^{(\ell)}(x)=u_{k,q+1}^{(\ell)}(x)\,\partial_p s_k^{(\ell)}(x).$$

根据 (4.2) 中块 $\mathcal{F}_{\ell,\ k}$ 的排列顺序，函数 $u_{k}^{(\ell)}\ ,\ q+1}$ 在列表中的位置早于 $u_k^{(\ell)}\ ,\ q}$，并且我们已经证明 $\partial_p s_k^{(\ell)}$ 是较早列表元素的多项式。因此，右侧是关于 $(x,\ f_1,\ \ldots\ldots,\ f_i)$ 的多项式，其次数至多为

$$1+D_{\ell-1}=D_{\ell-1}+1\leq D_\ell.$$

最后，利用 (4.5)，我们得到

$$\partial_p u_{k,r}^{(\ell)}(x)=\big(a_0+a_1u_{k,r}^{(\ell)}(x)+a_2(u_{k,r}^{(\ell)}(x))^2\big)\,\partial_p s_k^{(\ell)}(x),$$

它是列表元素 $u_{k\ r}^{(\ell)}\ ,$ 的一个次数为 2 的多项式因子，乘以前面列表元素中的一个次数至多为 $D_{\ell-1}$的多项式，因此在 $(x,\ f_1,\ \ldots,\ f_i)$ 中的总次数至多为 $D_{\ell-1}+2=D_\ell$。这就完成了证明的归纳步骤，因此该陈述对所有 $\ell\leq L$ 均成立。

取 $\alpha=D_L=1+2L$，则定义 4.1 的 Pfaffian链条件得到满足，因为对于每个列表元素 f_i和每个坐标指标 p，我们都已将 $\partial_p f_i$ 表示为 $(x,\ f_1,\ \ldots,\ f_i)$ 中的多项式：事实上，(4.3) 仅使用 ℓ之前各层中的函数来表示 $\partial_p s_k^{(\ell)}$ ，这些函数在列表中出现得更早；而 (4.4)–(4.5) 则使用 $s_k^{(\ell)}$ 以及满足 $q'>q$的函数 $u_k^{(\ell)}\ ,\ q'}$ 来表示 $\partial_p u_k^{(\ell)}\ ,\ q}$ ，它们都在同一块中更早出现。这就完成了子引理的证明。•

尚需将网络输出表示为列表中的多项式。由 (2.2),

$$F(x)=c_0+c^\top h^{(L)}(x)=c_0+\sum_{k=1}^{n_L}c_k\,h_k^{(L)}(x)=c_0+\sum_{k=1}^{n_L}c_k\,u_{k,0}^{(L)}(x),$$

它是列表元素中总次数为 1 的多项式。因此，在定义 4.1 的意义下可取 $\beta=1$ 。将长度 R、次数界 $\alpha=1+2L$ 与 $\beta=1$ 结合起来，即得格式 (4.1)。□

命题*3.1*的证明。 假设 $d=1$ ，并设 $I\subset\mathbb{R}$ 为一个开区间，在该区间上 F 是实解析的。根据带 $d=1$的命题 4.5， F 在 I 上的限制是格式为

$$(1,R,\alpha,\beta)=\Big(1,\ (r+2)\sum_{\ell=1}^{L}n_\ell,\ 1+2L,\ 1\Big),$$

因此 R 与 (3.1) 一致。若 $F\not\equiv 0$ 在 I上，考虑零点集

$$Z:=\{x\in I:\ F(x)=0\}.$$

由于 F 在 I上是 Pfaffian 的，集合 Z 在 I 上是半 Pfaffian 的，且由单个普法夫等式描述。此外，因为 F 是实解析的且不恒为零， Z 是有限的，因此

$$\mathrm{Zeros}(F;I)=|Z|=b_0(Z)=\mathrm{Betti}(Z).$$

将定理 4.4 应用于 Z （其中 $d=1$ 和 $s=1$），得到

$$\mathrm{Zeros}(F;I)\leq 2^{\frac{R(R-1)}{2}}\ C_I\ \Big(\beta+\min\{1,R\}\alpha\Big)^{1+R}=2^{\frac{R(R-1)}{2}}\ C_I\ (\alpha+\beta)^{1+R}.$$

代入 $\alpha = 1 + 2L$ 和 $\beta = 1$，得到

$$\mathrm{Zeros}(F; I) \leq 2^{\frac{R(R-1)}{2}}\ C_I\ \big(2 + 2L\big)^{1+R} = 2^{\frac{R(R+1)}{2}+1}\ C_I\ (1 + L)^{1+R}.$$

将常数因子 2 吸收到 C_I 中，得到

$$\mathrm{Zeros}(F; I)\ \leq\ 2^{\frac{R(R+1)}{2}}\, C_I\, (1 + L)^{R+1}.$$

若 $F \equiv 0$ 在 I 上，则 $\mathcal{S}_{\geq 0}(F) = I$，区间分量界显然成立。

现在，假设 $F \not\equiv 0$，并令 $z_1 < \cdots < z_N$ 为 I 中 F 的互异零点，其中 $N := \mathrm{Zeros}(F; I)$。于是 I 被划分为至多 $N + 1$ 个开区间

$$I \cap (-\infty, z_1), \quad (z_1, z_2), \quad \ldots, \quad (z_{N-1}, z_N), \quad I \cap (z_N, \infty).$$

在每个这样的子区间 F 上都没有零点，因此符号恒定；所以，限制在该子区间上的 $\{x \in I :\ F(x) \geq 0\}$ 要么为空集，要么就是整个子区间。因此，$\mathcal{S}_{\geq 0}(F) = \{x \in I :\ F(x) \geq 0\}$ 是至多 $N + 1$ 个区间的并。将加法常数吸收到 C_I 中，并代入上述关于 N 的界，即得所述区间界。□

定理 *3.2* 的证明。 设 $V \subset \mathbb{R}^d$ 为一个开域，在其上 F 是实解析的。由命题 4.5， F is 在 V 上为 Pfaffian，且对于 mat

$$(d, R, \alpha, \beta) = \Big(d,\ (r + 2) \sum_{\ell=1}^{L} n_\ell,\ 1 + 2L,\ 1\Big).$$

根据定义 4.2，决策区域

$$\mathcal{S}_{\geq 0}(F) = \{x \in V :\ F(x) \geq 0\}$$

在 V 上是半 Pfaffian 的，由单个庞加莱不等式描述。将定理 4.4 应用于 $s = 1$ 可得

$$\mathrm{Betti}\big(\mathcal{S}_{\geq 0}(F)\big)\ \leq\ \mathcal{C}_V(d, 1, R, \alpha, \beta) = 2^{\frac{R(R-1)}{2}}\ C_V\ \Big(d\beta + \min\{d, R\}\alpha\Big)^{d+R}.$$

由于 $\beta = 1$ 且 $\alpha = 1 + 2L$，我们得到

$$\mathrm{Betti}\big(\mathcal{S}_{\geq 0}(F)\big) \leq 2^{\frac{R(R-1)}{2}}\ C_V\ \Big(d + \min\{d, R\}(1 + 2L)\Big)^{d+R} = \mathfrak{B}_V(d, R, L),$$

这即为所声称的界。特别地，连通分支的数目满足

$$b_0(\mathcal{S}_{\geq 0}(F)) \leq \mathrm{Betti}(\mathcal{S}_{\geq 0}(F)) \leq \mathfrak{B}_V(d, R, L).$$

□

定理 3.3 的证明依赖于一些预备结果，下面我们将予以证明。我们先介绍 Pfaffian 函数的以下闭包性质，这些性质可直接由定义 4.1 和链式法则推出。为完整起见，我们给出证明。

引理 4.6. 设 $V \subset \mathbb{R}^d$ 是开集。

> *(1)* 若 $G_1, \ldots, G_N$ 在 V 上是 *Pfaffian* 函数，则任何多项式组合 $P(G_1, \ldots, G_N)$ 在 V 上都是 *Pfaffian* 函数。*(2)* 若 G 在 V 上是 *Pfaffian* 函数，则每个偏导数 $\partial_{x_p} G$ 在 V 上都是 *Pfaffian* 函数；必要时，可通过添加有限个辅助 *Pfaffian* 函数来扩充底层 *Pfaffian* 链。

证明。 我们依次证明 (1) 和 (2)。

证明 (1)。选 取一个 Pfaffian 链 $(f_1, \ldots, f_R)$，定义在 V 上，使得每个 G_j 都可以写成

$$G_j(x) = Q_j\big(x, f_1(x), \ldots, f_R(x)\big).$$

若函数 G_j 是相对于不同 Pfaffian 链给出的，我们首先将这些链连接成一个单独的链，并相应地重写每个 G_j。然后， $P(G_1, \ldots, G_N)$ 是关于 $(x, f_1, \ldots, f_R)$ 的多项式，因此根据定义 4.1，它是 Pfaffian 函数。

（2）的证明。设 $(f_1, \dots, f_R)$ 是 V 上的一个 Pfaffian 链，且 Q 是一个多项式，使得 $G(x) = Q\big(x, f_1(x) \dots, f_R(x)\big)$。对其求导并应用链式法则，得到

$$\partial_{x_p} G = \frac{\partial Q}{\partial x_p}(x, f(x)) + \sum_{i=1}^{R} \frac{\partial Q}{\partial y_i}(x, f(x))\, \partial_{x_p} f_i(x), \qquad f(x) := \big(f_1(x), \dots, f_R(x)\big).$$

由 P 阶链性质，对每个 i，有

$$\partial_{x_p} f_i(x) = P_{i,p}\big(x, f_1(x), \dots, f_i(x)\big)$$

对于某个多项式 $P_{i,\,p}$。代入这些表达式可知，$\partial_{x_p} G$ 是 $(x,\ f_1,\ \dots,\ f_R)$ 的多项式，因此在 V 上是 Pfaffian 函数。如有需要，我们可以将该多项式表达式中出现的有限多个辅助 P 阶函数并入链中，从而扩充此链。□

Lemm引理 4.7。设 $V \subset \mathbb{R}^d$ 为开集，且 X、Y 是 V 上的实解析向量场，写作

$$X(z) = \sum_{a=1}^{d} X_a(z)\, \partial_{z_a}, \qquad Y(z) = \sum_{a=1}^{d} Y_a(z)\, \partial_{z_a},$$

其中每个系数函数 X_a、Y_a：$V \to \mathbb{R}$ 都是 V 上的 ***Pfaffian*** 函数。那么，李括号 $[X, Y]$ 的每个系数函数都是 V 上的 ***Pfaffian*** 函数。

证明。对于每个 $a \in \{1, \dots, d\}$，$[X, Y]$ 的第 a 个系数由下式给出

$$[X, Y]_a = \sum_{b=1}^{d} \Big(X_b\, \partial_{z_b} Y_a - Y_b\, \partial_{z_b} X_a \Big).$$

因此，$[X, Y]_a$ 是由 Pfaffian 函数 $\{X_b, Y_b\}_{b=1}^{d}$ 通过有限次求和与求积，以及偏导数 $\partial_{z_b} X_a$ 和 $\partial_{z_b} Y_a$ 得到的。根据引理 4.6，Pfaffian 函数在求和、求积和偏微分下封闭（必要时可扩充底层链）。因此，每个 $[X, Y]_a$ 都是 V 上的 Pfaffian 函数。□

定理 *3.3* 的证明。根据假设，对于每个 $i \in \{1, \dots, m\}$ 和 $p \in \{1, \dots, d\}$，系数函数$X_{i,\,p}$：$V \to \mathbb{R}$ 都是在 V 上由具有固定架构 $(d,\ L,\ n_1,\ \dots,\ n_L)$ 且激活函数为 $\sigma \in \mathcal{A}_{\mathrm{quad},\ r}$ 的前馈网络实现的，并且在 V 上实解析。对每个这样的网络，应用输入维度为 d 的命题 4.5，可知每个 $X_{i,\,p}$ 都是在 V 上的 Pfaffian 函数，其格式界仅取决于 $(d,\ L,\ n_1,\ \dots,\ n_L)$ 和 r，而与权重无关。

现在固定 $k \in \mathbb{N}$。注意，每个括号 $Y \in \mathcal{B}_k$ 都是通过对 $X_1,\ \dots,\ X_m$ 反复应用李括号运算至多 k 次得到的。因此，反复应用引理 4.7 可知，每个 $Y \in \mathcal{B}_k$ 的每个系数函数都是 V 上的 Pfaffian 函数，从而矩阵 $A_k(z)$ 的每个元素都是 V 上的 Pfaffian 函数。

现在固定 $\rho \in \{0, 1, \dots, d\}$，并将 M_1、…、$M_{s_{k,\,\rho}}$记为 A_k 的所有 $(\rho+1) \times (\rho+1)$ 子式的列表，其中

$$s_{k,\rho} := \binom{d}{\rho+1} \binom{|\mathcal{B}_k|}{\rho+1}.$$

每个子式 $M_i(z)$ 都是 $A_k(z)$ 的一个子矩阵的行列式，因此是 $A_k(z)$ 各项的多项式；根据引理 4.6(1)，每个 M_i 在 V 上都是 Pfaffian。

对于任意实矩阵 A，秩 $A \le \rho$ 当且仅当所有 $(\rho+1) \times (\rho+1)$ 子式均为零。将此应用于 $A = A_k(z)$，可得实解析描述

$$\mathcal{Z}_{k,\rho} = \{z \in V :\ \operatorname{rank} A_k(z) \le \rho\} = \bigcap_{i=1}^{s_{k,\rho}} \{z \in V :\ M_i(z) = 0\}.$$

因此 $\mathcal{Z}_{k,\ \rho}$ 是半 Pfaffian ffian 在 V 上，因为它表示为 $s_{k,\ \rho}$ 个 Pfaffian 等式。应用定理 4.4 可得 s

$$\begin{aligned}\mathrm{Betti}(\mathcal{Z}_{k,\rho}) \ &\leq \ \mathfrak{R}_V\Big(d, m, k, \rho;\, L, n_1, \ldots, n_L, r\Big) \\ &:= 2^{\frac{R_k(R_k-1)}{2}}\ (s_{k,\rho})^d\ C_V\ \Big(d\beta_k + \min\{d, R_k\}\alpha_k\Big)^{d+R_k},\end{aligned}$$

对于任意适用于子式 M_i 的统一 Pfaffian 格式界（$d,\ R_k,\ \alpha_k,\ \beta_k$），均成立，这证明了该结论。 □

参考文献

[1] A. Agrachev、D. Barilari 和 U. Boscain，《次黎曼几何综合导论》，剑桥高等数学研究丛书，剑桥大学出版社，2019 年。[2] M. Bianchini 和 F. Scarselli，《神经网络分类器的复杂性：浅层架构与深层架构的比较》，IEEE 神经网络与学习系统汇刊 25（2014），第 8 期，1553–1565。[3] E. Ergen 和 M. Grillo，《ReLU 神经网络的拓扑表达能力》，第三十七届学习理论年会，2024 年，第 1599–1642 页。[4] J. Freitag，《非 Pfaffian》，arXiv 预印本 arXiv:2109.09230（2021）。[5] A. Gabrielov 和 N. Vorobjov，《使用 Pfaffian 函数和诺特函数进行计算的复杂性》，《微分方程中的标准形、分岔与有限性问题》137（2004），211–250。[6] A. Gabrielov、N. Vorobjov 和 T. Zell，《半代数集与子 Pfaffian 集的贝蒂数》，《伦敦数学会会刊》69（2004），第 1 期，27–43。[7] W. H. Guss 和 R. Salakhutdinov，《利用代数拓扑刻画神经网络的容量》，arXiv 预印本 arXiv:1802.04443（2018）。[8] A. Hatcher，《代数拓扑》，剑桥大学出版社，剑桥，2002 年。[9] V. Jurdjevic，《几何控制理论》，剑桥大学出版社，1997 年。[10] A. G. Khovanskiĭ，《少项式》，第 88 卷，数学专著译丛，1991 年。[11] G. Montúfar、R. Pascanu、K. Cho 和 Y. Bengio，《深度神经网络的线性区域数量》，神经信息处理系统进展 27（2014）。[12] G. Naitzat、A. Zhitnikov 和 L. H. Lim，《深度神经网络的拓扑》，《机器学习研究期刊》21（2020），第 184 号，1–40。[13] M. Raghu、M. Maithra、B. Poole、J. Kleinberg、S. Ganguli 和 J. Sohl Dickstein，《深度神经网络的表达能力》，国际机器学习会议，2017 年，第 2847–2854 页。[14] P. Tabuada 和 B. Gharesifard，《通过非线性控制理论实现深度残差神经网络的通用逼近能力》，国际学习表征会议，2021 年。[15] ，《通过控制的视角看深度残差神经网络的通用逼近能力》，IEEE 自动控制汇刊 68（2022），第 6 期，2715–2728。[16] ，《“通过控制的视角看深度残差神经网络的通用逼近能力”的勘误》，IEEE 自动控制汇刊（2024），即将刊出。[17] M. Telgarsky，《神经网络中深度的益处》，学习理论会议，PMLR，2016 年，第 1517–1539 页。[18] T. Zell，《半 Pfaffian 集的贝蒂数》，《纯粹与应用代数杂志》139（1999），第 1–3 期，323–338。

加拿大安大略省金斯顿，女王大学数学与统计学系电子邮箱：bahman.gharesifard@queensu.ca网址：https://gharesifard.github.io