Neural Networks in Local Coordinates

Ahmed Abdeljawad
Johann Radon Institute for Computational and Applied Mathematics
Austrian Academy of Sciences
Altenberger Straße 69, 4040, Linz, Austria
Email: ahmed.abdeljawad@oeaw.ac.at

***Abstract*—We explore the ability of deep ReLU neural networks to realize functions on manifolds. By establishing appropriate assumptions, we ensure that the coordinate charts can be exactly represented without error. Locally, we construct networks characterizing tooth functions on coordinate neighborhoods. To resolve mismatches arising from the manifold's complex structure, we further develop a global tooth function defined over the entire manifold, effectively represented by a ReLU neural network.**

***Index Terms*—Deep ReLU Neural Networks, Learning on Manifolds, Function Representation, Network Complexity.**

I. Introduction

Deep neural networks (DNNs) have emerged as powerful tools for tackling data-driven tasks across various scientific fields. Traditionally designed for inputs in vector spaces, DNNs leverage well-defined operations tailored to such structures [1], [6], [12], [13], [16]. However, with advancements in sensing technologies, the ability to work with manifold data has become increasingly important, as such data often arise directly or indirectly in real-world applications. Moreover, it is often observed that many datasets of interest inherently reside on low-dimensional manifolds within higher-dimensional ambient spaces, further complicating the learning process and emphasizing the need for methods that respect this underlying geometry.

While DNNs have demonstrated remarkable empirical success in modeling complex problems, the theoretical understanding of their effectiveness remains limited.

In this paper, we focus on the ability of DNNs to exactly construct functions on manifolds, with a particular emphasis on *tooth functions*, which play a critical role in various theoretical aspects of DNNs [3], [5], [20]. In our setting, and thanks to the assumptions made about the given manifold, we investigated the realization of coordinate charts using DNNs. Furthermore, while the tooth function is initially defined locally, it can result in mismatches between different regions. By leveraging ReLU DNNs, we successfully constructed a general tooth function defined over the entire manifold without any mismatches. More specifically, the results of this paper aim to bridge the so-called *Theory-to-Practice* gap in deep learning [4], [17]. Therefore, a follow-up paper addressing the Theory-to-Practice gap for ReLU DNNs with manifold-based data will be available soon.

To the best of our knowledge, the setting we consider has not been explored in prior work. Most recent studies focus on smooth manifolds [7], [8], [10], [14], [23] and the ability of DNNs to approximate functions defined on them. By contrast, we address *Piecewise Linear Manifolds*, demonstrating that for this specific structure, tooth functions can be represented exactly, without approximation, using ReLU DNNs.

II. Piecewise linear d-manifold

Manifolds are widely used as an input domain and has been extensively explored in many works when dealing with low-dimensional data. In this section, we briefly recall some preliminaries regarding *piecewise linear* manifolds.

Definition 1. *Let $U \subseteq \mathbb{R}^d, V \subseteq \mathbb{R}^m$ be open sets. A function $f : U \to V$ is said to be* piecewise affine-linear *(PL) if it is continuous and there is a locally-finite decomposition $U = \bigcup_{i\in I} K_i$ into connected, closed subsets $K_i \subseteq U$ with respect to which $f|_{K_i}$ is affine-linear.*

For a *PL homeomorphism* it is required that both maps, the bijection and its inverse, are piecewise linear. Note that, on a d-dimensional manifold (d-manifold) $\mathcal{M}$, for any $x \in \mathcal{M}$, there exist an open neighborhood $U \subset \mathcal{M}$ of x and a homeomorphism φ such that $\varphi : U \mapsto \varphi(U) \subseteq \mathbb{R}^d$. The pair (U, φ) is called a chart for $\mathcal{M}$ around x. A chart essentially defines a local coordinate system on $\mathcal{M}$. An atlas of $\mathcal{M}$ is a collection of charts (U_i, φ_i), $i \in I$, such that $\mathcal{M} = \bigcup_{i\in I} U_i$. Given $(U_i, \varphi_i)_{i\in I}$ an atlas, the transition functions are $\varphi_i \circ \varphi_j^{-1} : \varphi_j(U_i \cap U_j) \mapsto \varphi_i(U_i \cap U_j)$ with the obvious convention that we consider $\varphi_i \circ \varphi_j^{-1}$ if and only if $U_i \cap U_j \neq \emptyset$.

Definition 2. *A manifold $\mathcal{M}$ is called a piecewise linear manifold of dimension d (PL d-manifold) if it is equipped with an open covering $\mathcal{M} = \bigcup_{i\in I} U_i$ and coordinate charts $\varphi_i : U_i \to (V_i \subseteq \mathbb{R}^d)$ to open subsets of $\mathbb{R}^d$ for which*

$$\varphi_i \circ \varphi_j^{-1} : \varphi_j(U_i \cap U_j) \to \varphi_i(U_i \cap U_j)$$

are PL homeomorphisms.

In our paper, we consider a compact PL d-manifold $\mathcal{M}$ embedded in $\mathbb{R}^D$, where $D > d$. Intuitively, although $\mathcal{M}$ is a (hyper) surface in $\mathbb{R}^D$, locally it is identified with $\mathbb{R}^d$.

Definition 3 (Continuous functions on PL d-manifolds)**.** *Let $\mathcal{M}$ be a manifold. A function $f : \mathcal{M} \to \mathbb{R}$ is continuous* *if for any chart (U,φ), $f\circ\varphi^{-1}:\varphi(U)\to\mathbb{R}$ is continuous. The space $C(\mathcal{M})$ is the space of all continuous real-valued functions on $\mathcal{M}$, equipped with the supremum norm $\|\cdot\|_{C(\mathcal{M})}$. The definition of continuous functions is independent of the choice of the chart (U,φ).*

III. ReLU Neural network in local coordinate

We describe the architecture of deep ReLU neural networks which we use throughout this paper. We will be concerned with target classes related to ReLU neural networks. Here we use bold symbols for vectors $\boldsymbol{x}\in\mathbb{R}^n$, $n\in\mathbb{N}$, and normal symbols for $x\in\mathcal{M}$. Let $\varrho:\mathbb{R}\to\mathbb{R}$, such that $\varrho(\boldsymbol{t})=\max\{0,\boldsymbol{t}\}$ be the *ReLU activation function.* Given a *depth* $L\in\mathbb{N}$, an *architecture* $(N_0,N_1,\ldots,N_L)\in\mathbb{N}^{L+1}$, and *neural network coefficients* $\Phi=\left((W^i,b^i)\right)_{i=1}^L\in\bigtimes_{i=1}^L\left(\mathbb{R}^{N_i\times N_{i-1}}\times\mathbb{R}^{N_i}\right)$, we define their *realization* $\mathbf{R}(\Phi)\in C(\mathbb{R}^{N_0},\mathbb{R}^{N_L})$ as $\mathbf{R}(\Phi)(\boldsymbol{x})=\boldsymbol{x}^L$, where $\boldsymbol{x}^0=\boldsymbol{x}\in\mathbb{R}^{N_0},\boldsymbol{x}^i=\varrho\left(W^i\boldsymbol{x}^{i-1}+b^i\right)\in\mathbb{R}^{N_i}$ for $i\in[L-1]$, and $\boldsymbol{x}^L=W^L\boldsymbol{x}^{L-1}+b^L\in\mathbb{R}^{N_L}$, with ϱ applied componentwise. Let $\mathcal{H}_{(N_0,\ldots,N_L),C}$ be the class of feed-forward neural networks with the architecture $\mathcal{A}=(N_0,\ldots,N_L)$ and a uniform coefficients upper bound $C>0$ as follows:

$$\mathcal{H}_{\mathcal{A};C}:=\Big\{\mathbf{R}(\Phi):\Phi\in\bigtimes_{i=1}^{L}\left(\mathbb{R}^{N_i\times N_{i-1}}\times\mathbb{R}^{N_i}\right)\mid\|\Phi\|_{\ell^\infty}\le C\Big\},$$

where $\|\Phi\|_{\ell^\infty}:=\max_{i=1,\ldots,L}\max\{\|W^i\|_{\ell^\infty},\|b^i\|_{\ell^\infty}\}$, such that $\|W^i\|_{\ell^\infty}:=\max_{k,j}|W^i_{k,j}|$ and $\|b^i\|_{\ell^\infty}:=\max_k|b^i_k|$.

A. ReLU hat function on PL d-manifolds

Let $d,D\in\mathbb{N}$, $R>0$ and define

$$\Lambda_R:\quad\mathbb{R}\times\mathbb{R}\to(-\infty,1],\quad(\boldsymbol{s},\boldsymbol{t})\mapsto1-R\cdot|\boldsymbol{t}-\boldsymbol{s}|\qquad(1)$$

furthermore using the ReLU activation function, we have $\Delta_R:\mathbb{R}^D\times\mathbb{R}^D\to(-\infty,1],(\boldsymbol{\xi},\boldsymbol{x})\mapsto\left(\sum_{i=1}^D\Lambda_R(\boldsymbol{\xi}_i,\boldsymbol{x}_i)\right)-(D-1)$ and

$$\vartheta_R:\mathbb{R}^D\times\mathbb{R}^D\to[0,1],(\boldsymbol{\xi},\boldsymbol{x})\mapsto\varrho(\Delta_R(\boldsymbol{\xi},\boldsymbol{x}))\qquad(2)$$

Our aim is to extend the previous construction to $\mathcal{M}$. First, we construct a finite number of coordinates neighborhoods that cover $\mathcal{M}$. Let $\mathcal{B}(z,r)$ denotes the open ℓ^∞ ball in $\mathbb{R}^D$ with center $z\in\mathcal{M}$ and radius $r>0$: $\mathcal{B}(z,r):=\{\boldsymbol{x}\in\mathbb{R}^D$ such that $\sup_{j\in\{1,\ldots,D\}}|\boldsymbol{x}_j-\boldsymbol{z}_j|<r\}$, where $(\boldsymbol{z}_1,\ldots,\boldsymbol{z}_D)$ is the coordinate of z in $\mathbb{R}^D$. Here, we assume that $\mathcal{M}\subseteq\mathcal{B}(0,1)=\mathcal{B}(0_{\mathbb{R}^D},1)$. The properties of $\mathcal{M}$ imply the existence of a fixed number $\mathsf{C}_\mathcal{M}\in\mathbb{N}$, a finite collection of points $c_i\in\mathcal{M}$ for $i=1,\ldots,\mathsf{C}_\mathcal{M}$, and $r>0$, such that $\mathcal{M}\subseteq\bigcup_i\mathcal{B}_i(r)\equiv\mathcal{B}_i(r)$.

We proceed under the following list of assumptions.

Assumption 1. *Let $0<r<\infty$ such that there exists $c_i\in\mathcal{M}$, $i\in\mathcal{I}_\mathcal{M}=\{1,\ldots,\mathsf{C}_\mathcal{M}\}$, with the following properties:*

1) *$\mathcal{M}\subseteq\bigcup_{i=1}^{\mathsf{C}_\mathcal{M}}\mathcal{B}_i(r)$;*
2) *$U_i=\mathcal{B}_i(r)\cap\mathcal{M}$ is a convex set for any $i\in\mathcal{I}_\mathcal{M}$;*
3) *for any $i\neq j$ where $i,j\in\{1,\ldots,\mathsf{C}_\mathcal{M}\}$, we have*

$$U_i{}^{\circ}\cap U_j{}^{\circ}=\emptyset.\qquad(3)$$

We define our coordinate neighborhood U_i as the intersection $\mathcal{B}_i(r)\cap\mathcal{M}$ for any $i\in\{1,\ldots,\mathsf{C}_\mathcal{M}\}$. Let the tangent space of $\mathcal{M}$ at c_i denoted by $T_{c_i}(\mathcal{M})=\mathrm{span}(\mathbf{v}_{i1},\ldots,\mathbf{v}_{id})$, where $\{\mathbf{v}_{i1},\ldots,\mathbf{v}_{id}\}$ form an orthonormal basis, then $V_i=[\mathbf{v}_{i1},\ldots,\mathbf{v}_{id}]$ is a *projection matrix* belongs to $\mathbb{R}^{D\times d}$. Furthermore, V_i is a semi-orthogonal matrix, that is, $V_i^TV_i=I_d$, where I_d is the identity matrix of size d. For any $x\in U_i$, we define the projection ϕ_i as a coordinate system, $i=1,\ldots,\mathsf{C}_\mathcal{M}$, as follows:

$$\phi_i(x)=R_d(o_i)(V_i^T(x-c_i)+b_i),\qquad(4)$$

where $R_d(o_i)$ is a $d\times d$ rotation matrix where o_i is the angle of rotation and $b_i\in\mathbb{R}^d$ is a translation vector, such that there exists $r>0$ (from Assumption 1) where

$$[-r/2,r/2]^d\subseteq\phi_i(U_i)=\phi_i(\mathcal{B}_i(r)\cap\mathcal{M})\subseteq[-r\sqrt{D},r\sqrt{D}]^d.\qquad(5)$$

Consequently

$$r\sqrt{d}\le\mathrm{diam}(\phi_i(U_i))\le2r\sqrt{dD}.\qquad(6)$$

Note that (compared to [8], [9], [19], [21], [22]) in our construction of ϕ_i, we use a different factor, that is, the rotation matrix $R_d(o_i)$ instead of a scaling factor. The fact that $\mathcal{M}$ is a connected manifold insures that the center c_i of any coordinate neighborhood U_i is at least connected in U_i to the boundary of the ℓ^∞ ball $\mathcal{B}_i(r)$ and at most it belongs in the diagonal of $\mathcal{B}_i(r)$. The right-hand interval $[-r\sqrt{D},r\sqrt{D}]^d$ in (5) contains the maximum projection of U_i with respect to ϕ_i. The maximum projection refers to the diagonal of the ℓ^∞ ball, hence $\sqrt{D}$ appears since we use the ℓ^∞ norm in $\mathbb{R}^D$. While $[-r/2,r/2]^d$ refers to the minimum projection of U_i with respect to ϕ_i.

Each ϕ_i is a linear function which can be constructed by a single layer ReLU network. Therefore, let $r>0$ satisfies Assumption 1 and $\{(U_i,\phi_i)\}_{i=1}^{\mathsf{C}_\mathcal{M}}$ be an atlas on $\mathcal{M}$.

At this point, our paper will continue with an atlas $\{(U_i,\phi_i)\}_{i=1}^{\mathsf{C}_\mathcal{M}}$ that fulfills Assumption 1 such that ϕ_i is given by (4) and satisfies (5) for any $i\in\{1,\ldots,\mathsf{C}_\mathcal{M}\}$.

In order to construct a tooth function on $\mathcal{M}$, we use similar techniques to the construction of a bump function on manifolds cf., [24, Chapter 13]. We define a tooth function on $\phi_i(U_i)$, $i\in\{1,\ldots,\mathsf{C}_\mathcal{M}\}$, by first letting c_i be the center of U_i and $\phi_i(c_i)=\boldsymbol{c}_i\in[-r/2,r/2]^d$, where $\Phi_i:U_i\times U_i\to\phi_i(U_i)\times\phi_i(U_i)$, $(c_i,x)\mapsto(\phi_i(c_i),\phi_i(x))$ $\Phi_i^{-1}:\phi_i(U_i)\times\phi_i(U_i)\to U_i\times U_i$, $(\boldsymbol{c}_i,\boldsymbol{x})\mapsto(c_i,\phi_i^{-1}(\boldsymbol{x}))$. Hence, we define a tooth function on $\phi_i(U_i)\times\phi_i(U_i)\subset[-r\sqrt{D},r\sqrt{D}]^d\times[-r\sqrt{D},r\sqrt{D}]^d$

$$\begin{aligned}\Gamma_R^i:&[-r\sqrt{D},r\sqrt{D}]^d\times[-r\sqrt{D},r\sqrt{D}]^d\to[0,1],\\&(\boldsymbol{c}_i,\boldsymbol{x})\mapsto\vartheta_R\circ\Phi_i^{-1}(\boldsymbol{c}_i,\boldsymbol{x}).\end{aligned}\qquad(7)$$

Consequently, a tooth function on $\mathcal{M}$ can be defined as

$$\begin{aligned}\Gamma_R^i\circ\Phi_i:&U_i\times U_i\to[0,1],\\&(c_i,x)\mapsto\Gamma_R^i\circ\Phi_i(c_i,x)=\left(\vartheta_R\circ\Phi_i^{-1}\right)\big|_{\Phi_i(c_i,x)}.\end{aligned}$$

Defining functions on manifolds in the previous way is already known for deep learning [11]. The construction of a geodesic distance on $\mathcal{M}$ throw ReLU networks is not a part of the scope of this paper. Instead we use L^∞ distance from the ambient space $\mathbb{R}^D$ in the construction of the tooth function. Next, we derive some properties of Γ_R^i.

Theorem 4. *Let $r, R > 0$, $\mathcal{M}$ be a manifold and $\{(U_i, \phi_i)\}_{i=1}^{\mathsf{C}_\mathcal{M}}$ be an atlas on $\mathcal{M}$. Furthermore, let $c_i \in \mathcal{M}$ be the center of U_i such that $\phi_i(c_i) = \boldsymbol{c}_i \in [-r/2, r/2]^d$, and $i \in \{1, \dots, \mathsf{C}_\mathcal{M}\}$. Then the tooth function Γ_R^i from (7) satisfies the following:*

$$\operatorname{supp} \Gamma_R^i(\boldsymbol{c}_i, \cdot) \subset \phi_i\left(\mathcal{B}_i(R^{-1})\right) \cap \phi_i\left(U_i\right)$$

and let $\tau = (2RD)^{-1}$, we have

$$\frac{1}{2}\left(\frac{2\sqrt{2}}{3}\right)^{d(d-1)/p} \cdot \min(\tau, r)^{d/p} \leq \|\Gamma_R^i(\boldsymbol{c}_i, \cdot)\|_{L^p([-r\sqrt{D}, r\sqrt{D}]^d)} \leq \left(2\sqrt{D} \cdot \min(R^{-1}, r)\right)^{d/p}.$$

Proof. We use the early constructed atlas $\{(U_i, \phi_i)\}_{i=1}^{\mathsf{C}_\mathcal{M}}$ in Section III-A. For a fixed $c_i \in U_i$, in view of (2) and (7) the tooth function $\Gamma_R^i(\boldsymbol{c}_i, \boldsymbol{x}) \neq 0$ only if $\Delta_R(\phi_i^{-1}(\boldsymbol{c_i}), \phi_i^{-1}(\boldsymbol{x})) > 0$. Consequently,

$$\sum_{j=1}^{D} \Lambda_R(\phi_i^{-1}(\boldsymbol{c_i})_j, \phi_i^{-1}(\boldsymbol{x})_j) > D - 1.$$

Since $\Lambda_R(\phi_i^{-1}(\boldsymbol{c_i})_j, \phi_i^{-1}(\boldsymbol{x})_j) \in (-\infty, 1]$, we need that $\Lambda_R(\phi_i^{-1}(\boldsymbol{c_i})_j, \phi_i^{-1}(\boldsymbol{x})_j) > 0$ for all $j \in \{1, \dots, D\}$ which is (in view of (1)) possible if $|\phi_i^{-1}(\boldsymbol{c_i})_j - \phi_i^{-1}(\boldsymbol{x})_j| \leq \frac{1}{R}$ for any $j \in \{1, \dots, D\}$, hence

$$\boldsymbol{x} \in \phi_i\left(\mathcal{B}(\phi_i^{-1}(\boldsymbol{c}_i), R^{-1}) \cap U_i\right) = \phi_i\left(\mathcal{B}_i(R^{-1})\right) \cap \phi_i\left(U_i\right),$$

last equality holds true since ϕ_i is injective, for any $i \in \{1, \dots, \mathsf{C}_\mathcal{M}\}$. Concerning the lower and upper bounds for $\Gamma_R^i(\boldsymbol{c}_i, \cdot)$ we recall that by assumption

$$\operatorname{diam}\left(\phi_i(U_i)\right) = \operatorname{diam}\left(\phi_i\left(\mathcal{B}_i(r) \cap \mathcal{M}\right)\right) \leq 2r\sqrt{dD},$$

(see (5) and its consequence), hence

$$\begin{aligned}\phi_i\left(\mathcal{B}_i(R^{-1})\right) \cap \phi_i\left(U_i\right) &= \phi_i\left(\mathcal{B}_i(R^{-1})\right) \cap \phi_i\left(\mathcal{B}_i(r) \cap \mathcal{M}\right) \\ &= \phi_i\left(\mathcal{B}_i(R^{-1}) \cap \mathcal{B}_i(r) \cap \mathcal{M}\right) \\ &= \phi_i\left(\mathcal{B}_i(\min(R^{-1}, r)) \cap \mathcal{M}\right),\end{aligned}$$

the second equality holds true thanks to the injectivity of ϕ_i. Then, we conclude that

$$\operatorname{diam}\left(\phi_i\left(\mathcal{B}_i(R^{-1})\right) \cap \phi_i\left(U_i\right)\right) \leq 2 \cdot \min(R^{-1}, r) \cdot \sqrt{dD}.$$

Without loss of generality, the set $\phi_i\left(\mathcal{B}_i(R^{-1})\right) \cap \phi_i\left(U_i\right)$ can be thought as the product of intervals of the form $(\boldsymbol{a}_i, \boldsymbol{b}_i)$ such that $0 < \boldsymbol{a}_i < \boldsymbol{b}_i < 1$ (see (5)) where $i \in \{1, \dots, d\}$. We can bound the Lebesgue measure λ_d on $\mathbb{R}^d$ of a bounded measurable set $E = \prod_{i=1}^d (a_i, b_i) \subset \mathbb{R}^d$ as follows:

$$\begin{aligned}\lambda_d(E) &= \prod_{i=1}^{d}(b_i - a_i) = \left(\prod_{i=1}^{d}(\boldsymbol{b}_i - \boldsymbol{a}_i)^2\right)^{1/2} \\ &\leq \left(\frac{1}{d}\sum_{i=1}^{d}(\boldsymbol{b}_i - \boldsymbol{a}_i)^2\right)^{d/2} = \left(\frac{1}{d}\right)^{d/2} \operatorname{diam}(E)^d.\end{aligned} \tag{8}$$

Hence, for any $\boldsymbol{x} \in \phi_i\left(\mathcal{B}_i(R^{-1})\right) \cap \phi_i\left(U_i\right)$, we get

$$\begin{aligned}&\left\|\Gamma_R^i(\boldsymbol{c}_i, \cdot)\right\|_{L^p(([-r\sqrt{D}, r\sqrt{D}]^d)} \\ &\leq \lambda_d\left(\phi_i\left(\mathcal{B}_i(R^{-1})\right) \cap \phi_i\left(U_i\right)\right)^{\frac{1}{p}} \leq \left(2 \cdot \min(R^{-1}, r)\sqrt{D}\right)^{\frac{d}{p}}.\end{aligned}$$

For the lower bound, we use a similar idea from [5, Proof of Lemma 2.3]. That is, let $\boldsymbol{x} \in \phi_i\left(\mathcal{B}_i(\tau)\right) \cap \phi_i\left(U_i\right)$, then $\phi_i^{-1}(\boldsymbol{x}) \in \mathcal{B}_i(\tau) \cap U_i$ which implies that for any $j \in \{1, \dots, D\}$ we have

$$\Lambda_R\left(c_{ij}, \phi_i^{-1}(\boldsymbol{x})_j\right) = 1 - R\left|\phi_i^{-1}(\boldsymbol{x}))_j - c_{ij}\right| \geq 1 - \frac{1}{2D}.$$

Consequently, we get $\Delta_R\left(c_i, \phi_i^{-1}(\boldsymbol{x})\right) = \left(\sum_{j=1}^{D} \Lambda_R\left(c_{ij}, \phi_i^{-1}(\boldsymbol{x})_j\right)\right) - (D-1) \geq \frac{1}{2}$, and hence

$$\Gamma_R^i(\boldsymbol{c}_i, \boldsymbol{x}) \geq \frac{1}{2} \text{ for any } \boldsymbol{x} \in \phi_i\left(\mathcal{B}_i(\tau)\right) \cap \phi_i\left(U_i\right). \tag{9}$$

The fact that $\phi_i\left(\mathcal{B}_i(\tau)\right) \cap \phi_i\left(U_i\right) = \phi_i\left(\mathcal{B}_i(\tau) \cap \mathcal{B}_i(r) \cap \mathcal{M}\right) = \phi_i\left(\mathcal{B}_i(\min\left(\tau, r\right) \cap \mathcal{M}\right)$ implies that (in view of (6)) for any $j \in \{1, \dots, d\}$ the sides s_j of $\phi_i\left(\mathcal{B}_i(\tau)\right) \cap \phi_i\left(U_i\right)$ satisfy

$$s_j \in [\min\left(\tau, r\right), 2 \cdot \min\left(\tau, r\right) \cdot \sqrt{D}]. \tag{10}$$

Furthermore, in our case $\min\left(\tau, r\right) \leq s_i \leq 2\sqrt{D} \cdot \min\left(\tau, r\right)$, hence, let $M = 2\sqrt{D} \cdot \min\left(\tau, r\right)$ and $m = \min\left(\tau, r\right)$ it follows that

$$\frac{(M+m)^2}{4Mm} \leq \frac{\left(M + m \cdot \sqrt{D}\right)^2}{4Mm} = \frac{9}{8}. \tag{11}$$

In view of (5), (8), [15, Theorem A] (10) and (11), we get $\lambda_d\left(\phi_i\left(\mathcal{B}_i(\tau)\right) \cap \phi_i\left(U_i\right)\right) \geq \left(\left(\frac{9}{8}\right)^{-\frac{d-1}{2}} \cdot \frac{1}{d} \cdot \sum_{j=1}^{d} s_j\right)^d \geq \left(\left(\frac{8}{9}\right)^{\frac{d-1}{2}} \cdot \frac{1}{d} \cdot \min(\tau, r) \cdot d\right)^d \geq \left(\frac{2\sqrt{2}}{3}\right)^{d(d-1)} \cdot \min(\tau, r)^d$. Overall, from (5), using the previous inequality and (9), we obtain the lower bound

$$\begin{aligned}\left\|\Gamma_R^i(\boldsymbol{c}_i, \cdot)\right\|_{L^p([-r\sqrt{D}, r\sqrt{D}]^d)} &\geq \frac{1}{2}\lambda_d\left(\phi_i\left(\mathcal{B}_i(\tau)\right) \cap \phi_i\left(U_i\right)\right)^{\frac{1}{p}} \\ &\geq \frac{1}{2}\left(\frac{2\sqrt{2}}{3}\right)^{d(d-1)/p} \cdot \min(\tau, r)^{\frac{d}{p}}.\end{aligned}$$

$\square$

In order to avoid any projection of a given x using unmatched charts, we need to construct an indicator function. For example if there exists $y \in \mathcal{M} \setminus U_i$ such that the orthogonal projection of y on $T_{c_i}\mathcal{M}$ is the same as $\phi_i(c_i)$ then $\Gamma_R^i \circ \Phi_i(c_i, y) = 1$. This case should be eliminated, that is, for $i \in \{1, \dots, \mathsf{C}_\mathcal{M}\}$ we would like to get $\Gamma_R^i \circ \Phi_i(c_i, y) = 0 \quad$ if $y \in \mathcal{M} \setminus U_i$. Hence, we construct a ReLU network $\mathbb{1}_{r,\delta}^i$ that satisfies:

$$\begin{cases} \mathbb{1}_{r,\delta}^i(x) = 1 & \text{if } x \in \mathcal{B}_i(r-\delta) \cap U_i, \\ 0 < \mathbb{1}_{r,\delta}^i(x) < 1 & \text{if } x \in \mathring{U}_i \setminus \mathcal{B}_i(r-\delta), \\ \mathbb{1}_{r,\delta}^i(x) = 0 & \text{if } x \notin \mathring{U}_i, \end{cases} \tag{12}$$

which can be considered as indicator function in our setting. We refer the reader to Section III-B for more details about the existence and the construction of the indicator function $\mathbb{1}^i_{r,\delta}$ using ReLU networks. Moreover, in view of Theorem 4 for $p=\infty$ and (12), if $x \notin U_i$ then $\max(0, \Gamma^i_{1/r} \circ \Phi_i(c_i,x) + \mathbb{1}^i_{r,\delta}(x) - 1) = \varrho(\Gamma^i_{1/r} \circ \Phi_i(c_i,x) - 1) = 0$. Furthermore, if $x \in \mathcal{B}_i(r-\delta) \cap U_i$, we have $\mathbb{1}^i_{r,\delta}(x) = 1$, which implies that $\max(0, \Gamma^i_{1/r} \circ \Phi_i(c_i,x) + \mathbb{1}^i_{r,\delta}(x) - 1) = \Gamma^i_{1/r} \circ \Phi_i(c_i,x)$. Consequently, the local tooth function (without mismatching) can be constructed as follows: $\Lambda^i_{r,\delta}(x) := \varrho(\Gamma^i_{1/r} \circ \Phi_i(c_i,x) + \mathbb{1}^i_{r,\delta}(x) - 1) = \Gamma^i_{1/r} \circ \Phi_i(c_i,x)$ if $x \in \mathcal{B}_i(r-\delta) \cap U_i$, and 0 if $x \notin U_i$. Finally, the global tooth function is defined as a summation of all local teeth functions, i.e., $\Lambda_{r,\delta}(x) = \sum_{i=1}^{\mathsf{C}_\mathcal{M}} \Lambda^i_{r,\delta}(x)$.

B. ReLU indicator function on PL manifolds

Mainly, the indicator function $\mathbb{1}_{U_i}$ of an input $x \in \mathcal{M}$ determines the chart that x belongs to, for any $i \in \{1, \ldots, \mathsf{C}_\mathcal{M}\}$, such that $U_i = \mathcal{B}_i(r) \cap \mathcal{M}$ satisfies Assumption 1. We use an indicator function in order to avoid any non proper coordinates neighborhood assignment $\phi_i(y)$ if $y \notin U_i$ where $i \in \{1, \ldots, \mathsf{C}_\mathcal{M}\}$.

Lemma 5. *Let $0 < \delta \leq r$, $d << D \in \mathbb{N}$ and $\tilde{N}_j \in \mathbb{N}$ such that $\tilde{N}_j \leq 3D$ where $j \in \{1, \ldots, D-1\}$. Let $\mathcal{M}$ be a d-dimensional PL manifold embedded in $\mathbb{R}^D$ and $\{(U_i, \phi_i)\}_{i=1}^{\mathsf{C}_\mathcal{M}}$ be an atlas on $\mathcal{M}$. Furthermore, let $c_i \in \mathcal{M}$ be the center of U_i and $\mathbb{1}^i_{r,\delta}$ defined as*

$$\mathbb{1}^i_{r,\delta}(x) = \begin{cases} 1 & \text{if } x \in \mathcal{B}_i(r-\delta) \cap U_i, \\ 1 - \frac{1}{\delta}(x - r + \delta) & \text{if } x \in U_i \setminus \mathcal{B}_i(r-\delta) \cap \mathcal{M} \\ 0 & \text{if } x \notin U_i. \end{cases}$$

then

$$\mathbb{1}^i_{r,\delta} \in \mathcal{H}_{(D, \tilde{N}_1, \ldots, \tilde{N}_{D-1}, 1, 1, 1),\, 2/\delta}.$$

Proof. In order to get a ReLU implementable indicator function, we decompose $\mathbb{1}_{U_i}$ in two functions, namely $\mathbb{1}_{[0,r]} \circ \mathcal{L}^\infty_i(x)$, such that $\mathbb{1}_{[0,r]}$ defined on $\mathbb{R}$ as follows:

$$\mathbb{1}_{[0,r]}(t) := \begin{cases} 1 & \text{if } t \in [0,r] \\ 0 & \text{otherwise.} \end{cases}$$

Further, let the distance function $\mathcal{L}^\infty_i : \mathcal{M} \to \mathbb{R}_+$ be defined as $\mathcal{L}^\infty_i(x) := \max\left(|x_1 - c_{i1}|, \ldots, |x_D - c_{iD}|\right)$. The function $\mathcal{L}^\infty_i$ measures the distance between any given data x and the center c_i of U_i with respect to the ℓ^∞ norm, in view of the Euclidean coordinate system in $\mathbb{R}^D$. The fact that the maximum can be implemented using ReLU neural networks without error, helps in the implementation of $\mathcal{L}^\infty_i$. In fact, if $x_1, x_2 \in \mathbb{R}$, we have

$$\max(x_1, x_2) = x_1 + \varrho(x_2 - x_1) \in \mathcal{H}_{(2,3,1),1} \tag{13}$$

this implies that the maximum between two elements can be implemented by a single hidden layer ReLU network, since $x_1 = \varrho(x_1) - \varrho(-x_1)$.

In view of (13), for any $x \in \mathcal{M}$, we can get $\mathcal{L}^\infty_i(x)$ through a recursion argument. Indeed, in order to get the maximum of (x_1, x_2, x_3) it follows that we only need a ReLU network with two hidden layers each contains at most 9 neurons, that is, for any $x_1, x_2, x_3 \in \mathbb{R}$, we have $\max(x_1, x_2, x_3) = x_1 + \varrho(x_2 - x_1) + \varrho(x_3 - (x_1 + \varrho(x_2 - x_1)))$ assuming that our assumption holds true up to $k > 3$, that is, $\max(x_1, \ldots, x_k)$ can be represented as ReLU neural networks with $k-1$ hidden layers each layer contains at most $3k$ neurons. Now, in order to represent $\max(x_1, \ldots, x_{k+1})$ as a ReLU network, we use (13): $\max(x_1, \ldots, x_{k+1}) = \max(\max(x_1, \ldots, x_k), x_{k+1})$ which can be seen as parallelization of two ReLU networks. Then, there exists a ReLU network $\mathtt{L}^\infty_i$ such that $\mathtt{L}^\infty_i(x) = \max\left(|x_1 - c_{i1}|, \ldots, |x_D - c_{iD}|\right)$, using $D-1$ layers each contains at most $3D$ neurons. Hence, there exists $\tilde{C}_2 > 0$ such that

$$\mathtt{L}^\infty_i \in \mathcal{H}_{(D, \tilde{N}_1, \ldots, \tilde{N}_{D-1}, 1), \tilde{C}_2}, \quad \text{where } \tilde{N}_1, \ldots, \tilde{N}_{D-1} \leq 3D. \tag{14}$$

We use the following function to approximate $\mathbb{1}_{[0,r]}$:

$$\mathbb{1}_{r,\delta}(t) = \begin{cases} 1 & t \leq r - \delta, \\ 1 - \frac{1}{\delta}(t - r + \delta) & t \in [r-\delta, r], \\ 0 & t \geq r, \end{cases} \tag{15}$$

where δ is the "width" of the error region (which should be small enough). The function $\mathbb{1}_{r,\delta}$ can be realized through a ReLU network $\mathbb{1}_{r,\delta}(t) = \varrho\big(1 - \frac{1}{\delta}\,\varrho(t - r + \delta)\big)$. Then, there exists $\tilde{C}_1 > 0$ (depends on δ) such that

$$\mathbb{1}_{r,\delta} \in \mathcal{H}_{(1,1,1,1), \tilde{C}_1}. \tag{16}$$

We can represent the function $\mathbb{1}_{r,\delta} \circ \mathcal{L}^\infty_i(x)$ through a composition of two ReLU neural networks, using (14) and (16), i.e., $\mathbb{1}_{r,\delta} \circ \mathtt{L}^\infty_i$ without error. The previous function and its realization with a ReLU network satisfy:

$$\mathbb{1}_{r,\delta} \circ \mathcal{L}^\infty_i(x) = \begin{cases} 1 & \text{if } x \in \mathcal{B}_i(r-\delta) \cap \mathcal{M}, \\ 1 - \frac{1}{\delta}(t - r + \delta) & \text{if } x \in U_i \setminus \mathcal{B}_i(r-\delta) \cap \mathcal{M}, \\ 0 & \text{if } x \notin U_i. \end{cases}$$

In conclusion, it follows by a concatenation of networks, (see e.g., [2], [18]), that

$$\mathbb{1}_{r,\delta} \circ \mathtt{L}^\infty_i \in \mathcal{H}_{(D, \tilde{N}_1, \ldots, \tilde{N}_{D+1}, 1), \tilde{C}}, \tag{17}$$

where $\tilde{C} = 2/\delta$ and $\tilde{N}_1, \ldots, \tilde{N}_{D+1} \leq 3D$. In order to be more precise with respect to the requirements in [18, Lemma 2.5], We have $\mathbf{R}(\Phi^1) = \mathbb{1}_{r,\delta} \in \mathcal{H}_{(1,1,1,1), \tilde{C}_1}$ and $\mathbf{R}(\Phi^2) = \mathtt{L}^\infty_i \in \mathcal{H}_{(D, \tilde{N}_1, \ldots, \tilde{N}_{D-1}, 1), \tilde{C}_2}$ here $\tilde{C}_2 = 1$. Then, we get $\mathbf{R}(\Phi^1) \circ \mathbf{R}(\Phi^2) = \mathbb{1}_{r,\delta} \circ \mathtt{L}^\infty_i = \mathbb{1}^i_{r,\delta}$ such that $\mathbb{1}^i_{r,\delta} \in \mathcal{H}_{(D, \tilde{N}_1, \ldots, \tilde{N}_{D+1}, 1), \max(\tilde{C}_2, \tilde{C}_1(1 \times \tilde{C}_2 + 1))} \subset \mathcal{H}_{(D, \tilde{N}_1, \ldots, \tilde{N}_{D-1}, 1, 1, 1), 2/\delta}$, where $\tilde{N}_1, \ldots, \tilde{N}_{D-1} \leq 3D$. Obviously the coefficients bound $\tilde{C} = 2/\delta$ is independent on the dimensions of the ambient space $\mathbb{R}^D$ and $\mathcal{M}$. ☐

Acknowledgment

This work benefited significantly from constructive feedback of Prof. Philipp Grohs.

References

[1] A. Abdeljawad and T. Dittrich, "Weighted sobolev approximation rates for neural networks on unbounded domains," 2024. [Online]. Available: https://arxiv.org/abs/2411.04108

[2] A. Abdeljawad and P. Grohs, "Approximations with deep neural networks in Sobolev time-space," *Analysis and Applications*, vol. 20, no. 03, pp. 499–541, 2022-05. [Online]. Available: https://www.worldscientific.com/doi/10.1142/S0219530522500014

[3] ——, "Sampling complexity of deep approximation spaces," *Analysis and Applications*, vol. 23, no. 01, p. 1–30, May 2024. [Online]. Available: http://dx.doi.org/10.1142/S0219530524500234

[4] B. Adcock and N. Dexter, "The gap between theory and practice in function approximation with deep neural networks," *SIAM Journal on Mathematics of Data Science*, vol. 3, no. 2, p. 624–655, Jan. 2021. [Online]. Available: http://dx.doi.org/10.1137/20M131309X

[5] J. Berner, P. Grohs, and F. Voigtlaender, "Learning reLU networks to high uniform accuracy is intractable," in *The Eleventh International Conference on Learning Representations*, 2023. [Online]. Available: https://openreview.net/forum?id=nchvKfvNeX0

[6] H. Bölcskei, P. Grohs, G. Kutyniok, and P. Petersen, "Optimal approximation with sparsely connected deep neural networks," *SIAM Journal on Mathematics of Data Science*, vol. 1, no. 1, p. 8–45, Jan. 2019. [Online]. Available: http://dx.doi.org/10.1137/18M118709X

[7] M. Chen, H. Jiang, W. Liao, and T. Zhao, "Efficient approximation of deep relu networks for functions on low dimensional manifolds," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/file/fd95ec8df5dbeea25aa8e6c808bad583-Paper.pdf

[8] ——, "Nonparametric regression on low-dimensional manifolds using deep ReLU networks: Function approximation and statistical recovery," *Information and Inference: A Journal of the IMA*, vol. 11, no. 4, pp. 1203–1253, 2022.

[9] X. Cheng and A. Cloninger, "Classification logit two-sample testing by neural networks for differentiating near manifold densities," *IEEE Transactions on Information Theory*, vol. 68, no. 10, pp. 6631–6662, 2022.

[10] C. K. Chui and H. N. Mhaskar, "Deep nets for local manifold learning," *Frontiers in Applied Mathematics and Statistics*, vol. 4, May 2018. [Online]. Available: http://dx.doi.org/10.3389/fams.2018.00012

[11] ——, "Deep Nets for Local Manifold Learning," *Frontiers in Applied Mathematics and Statistics*, vol. 4, p. 12, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/fams.2018.00012/full

[12] I. Daubechies, R. DeVore, S. Foucart, B. Hanin, and G. Petrova, "Nonlinear approximation and (deep) ReLU networks," *Constructive Approximation*, vol. 55, no. 1, p. 127–172, Apr. 2021. [Online]. Available: http://dx.doi.org/10.1007/s00365-021-09548-z

[13] R. DeVore, B. Hanin, and G. Petrova, "Neural network approximation," *Acta Numerica*, vol. 30, p. 327–444, May 2021. [Online]. Available: http://dx.doi.org/10.1017/S0962492921000052

[14] Y. Fang, I. Ohn, V. Gupta, and L. Lin, "Intrinsic and extrinsic deep learning on manifolds," *Electronic Journal of Statistics*, vol. 18, no. 1, pp. 1160 – 1184, 2024. [Online]. Available: https://doi.org/10.1214/24-EJS2227

[15] J. I. Fujii, M. Fujii, M. Nakamura, J. Pečarić, and Y. Seo, "A reverse inequality for the weighted geometric mean due to Lawson–Lim," *Linear algebra and its applications*, vol. 427, no. 2-3, pp. 272–284, 2007.

[16] R. Gribonval, G. Kutyniok, M. Nielsen, and F. Voigtlaender, "Approximation spaces of deep neural networks," *Constructive Approximation*, vol. 55, no. 1, p. 259–367, May 2021. [Online]. Available: http://dx.doi.org/10.1007/s00365-021-09543-4

[17] P. Grohs and F. Voigtlaender, "Proof of the theory-to-practice gap in deep learning via sampling complexity bounds for neural network approximation spaces," *Foundations of Computational Mathematics*, vol. 24, no. 4, p. 1085–1143, Jul. 2023. [Online]. Available: http://dx.doi.org/10.1007/s10208-023-09607-w

[18] I. Gühring, G. Kutyniok, and P. Petersen, "Error bounds for approximations with deep ReLU neural networks in Ws,p norms," *Analysis and Applications*, vol. 18, no. 05, pp. 803–859, 2019.

[19] Y. Jiao, G. Shen, Y. Lin, and J. Huang, "Deep nonparametric regression on approximately low-dimensional manifolds," *arXiv:2104.06708*, 2021.

[20] Y. Li, P. Grohs, and P. Petersen, "The sampling complexity of learning invertible residual neural networks," 2024. [Online]. Available: https://arxiv.org/abs/2411.05453

[21] H. Liu, M. Chen, T. Zhao, and W. Liao, "Besov function approximation and binary classification on low-dimensional manifolds using convolutional residual networks," in *International Conference on Machine Learning*. PMLR, 2021, pp. 6770–6780.

[22] H. Liu, H. Yang, M. Chen, T. Zhao, and W. Liao, "Deep nonparametric estimation of operators between infinite dimensional spaces," *arXiv:2201.00217*, 2022.

[23] J. Schmidt-Hieber, "Deep relu network approximation of functions on a manifold," 2019. [Online]. Available: https://arxiv.org/abs/1908.00695

[24] L. W. Tu, *An Introduction to Manifolds*, 2nd ed., ser. Universitext. Springer, 2010-10.