Estimate of the Neural Net Dimension Using Algebraic Topology and Lie Theory*

Luciano Melodia[0000−0002−7584−7287]
and Richard Lenz[0000−0003−1551−4824]

Chair of Computer Science 6
Friedrich-Alexander University Erlangen-Nürnberg
91058 Erlangen, Deutschland
{luciano.melodia,richard.lenz}@fau.de

Abstract. In this paper we present an approach to determine the smallest possible number of perceptrons in a neural net in such a way that the topology of the input space can be learned sufficiently well. We introduce a general procedure based on persistent homology to investigate topological invariants of the manifold on which we suspect the data set. We specify the required dimensions precisely, assuming that there is a smooth manifold on or near which the data are located. Furthermore, we require that this space is connected and has a commutative group structure in the mathematical sense. These assumptions allow us to derive a decomposition of the underlying space whose topology is well known. We use the representatives of the k-dimensional homology groups from the persistence landscape to determine an integer dimension for this decomposition. This number is the dimension of the embedding that is capable of capturing the topology of the data manifold. We derive the theory and validate it experimentally on toy data sets.

Keywords: Embedding dimension · Parameterization · Persistent homology · Neural nets · Manifold learning.

1 Introduction

Since the development of deep neural nets, their parameterization, in particular the smallest possible number of perceptual units in a layer, has been studied. This number is of importance for auto-encoding tasks that require extrapolation or interpolation of data, such as blind source separation or super-resolution. If this number is overestimated, unnecessary resources are consumed during training. If the number is too small, no sufficiently good estimate can be given. To solve this problem, we make the following contribution in this paper:

- We study topological invariants using statistical summaries of persistent homology on a filtered simplicial complex on the data.
- Using the theory of Lie groups, we specify the smallest possible embedding dimension so that the neural net is able to estimate a projection onto a space with the determined invariants.

* The code can be found at: https://github.com/karhunenloeve/Irhe.

2 Related work

The related work is divided into two sections. In the first section we present chronologically the milestones in the consideration of neural nets from a differential geometric perspective. In the second section we present work on earlier attempts to determine their minimum embedding dimension.

Riemannian geometry in neural nets The manifold of a dense neural net is in most cases Euclidean. Very recently, there have been results in the direction of other manifolds on which neural nets can operate. The spherical neural nets operate on a manifold with commutative group structure [7,11].

A metric is required to describe the geometric properties of such a manifold. The change of coordinate systems and the learned metric tensor of a Riemann manifold during back propagation of neural nets has been investigated [15]. Further, it has been shown that dense neural nets can't approximate all functions arbitrarily precisely [17]. In response to this result, neural nets with residual connections have been examined. Residual nets add the output of one layer to a deeper one and therefore bypass some of the layers in between. This type of neural net is indeed an universal function approximator [21].

Embedding dimension of neural nets Cybenko showed that dense neural nets with a single hidden layer can approximate any continuous function with a bounded domain with arbitrary small error [8]. The upper bound of the embedding dimension for ReLU nets has recently been proven to be $(n+4)$ with n being the sample dimension [24]. Bartlett et al. show that a nearly-tight bound for the VC dimension of a neural net with W weights and L layers has $O(WL \log W)$ but $\Omega(WL \log W/L)$ VC dimensions [2].

The depth of a neural net has been determined to be a parameter depending on the moduli of continuity of a function to be approximated [21]. Inspired by the theory of finite difference methods, such a residual layer has been defined as a forward or backward difference operator on a partition of the layers. The state space dimension with residual connections is homeomorphic to $\mathbb{R}^{k \cdot n}$, where $n = \dim M$ and k is the number of times the difference operator was used [16]. Thus, such a layer is capable of embedding into $k \cdot n$ dimensional space. The view that a neural net operates on a smooth manifold inspired us to investigate the invariants of this manifold for the theory discussed in this paper.

To these fundamental results we contribute a topological approach to parameterization problems. Similar to Futagami et al. [12], we study the manifold on which we suspect the data. Our approach is applicable to topological spaces in general and is based on persistent homology. We are looking for the dimension of a space with the same topological properties as data indicates and of which we know that it can be approximated. We use the simplest decomposition into a product space of realaxes and 1-spheres to obtain the same homology groups as the filtration of the data has. Thus, we obtain a computable lower bound for the embedding dimension of any neural net with differentiable structure.

3 Smooth manifolds, Lie groups and persistent homology

The assumption that a set of points lies on a manifold is more accurate than handling points in Euclidean space. A *topological manifold* M is a *Hausdorff* space, which means that at any two points $x, y \in M$ there are always two open neighborhoods $U_x, U_y \subset M$, so that their intersection is empty. This behaviour provides the intuitive feeling of the position of points in space. The topology (M, τ) is a set system that describes the structure of a geometric object. Thereby M itself and $\emptyset$ must be contained in τ, and any arbitrary union of open sets from τ and any intersections from τ must be in τ. If this set system has a smaller set that generates the topology τ with the aid of finitely many unions, then this smaller set is a *basis* of the topology. We demand of a manifold that the basis is at most countable. At last, M is locally Euclidean, i.e. each point has a neighborhood which can be homeomorphically mapped to a subset of $\mathbb{R}^n$. The integer n is the dimension of M. Further, we demand some additional smooth structure.

3.1 Smooth manifolds

A smooth structure is a stronger condition than a topological manifold which can be described by a family of continuous functions. This assumption can be justified by the theorem of Stone-Weierstrass which proves that any continuous function can be approximated arbitrarily precisely by a smooth one [25]. Smooth manifolds are described by a collection of local coordinate charts $\varphi : U \to \varphi(U) \subseteq \mathbb{R}^n$, homeomorphic to $\mathbb{R}^n$ and an open neighborhood (U_i, φ_i) in M which is covered by this map. The family of such maps that cover M is the *atlas* $\mathcal{A} = (U_i, \varphi_i)_{i \in I}$ [20]. The functions $x_1, \cdots, x_n : U \to \mathbb{R}$ are called local coordinates $\varphi(p) = (x_1(p), \cdots, x_n(p))$. If the atlas is maximal with respect to inclusion, then it is a *differentiable structure*. Each atlas for M is included in a maximal atlas. Because of the differentiable structure, the maps in $\mathcal{A}$ are also compatible with all other maps of $\mathcal{A}_{\max}$. Thus, it is sufficient to describe smooth manifolds by an atlas that is not maximal.

As long as the activation functions are smooth and of the same differentiable structure, they are also compatible and the manifold does not change while propagating the data trough the layers. Only the coordinate representation changes.

Lie groups The study of group theory deals with symmetries which can be expressed algebraically. A pair $(M, \bullet)$, consisting of a map $\bullet : M \times M \to M$ is called group if the map is associative, $x \bullet (y \bullet z) = (x \bullet y) \bullet z$, has a neutral element so that $x \bullet e_M = x$ and has an inverse element to each element of the group, $x \bullet x^{-1} = e_M$. A Lie group is a smooth manifold which is equipped with a group structure such that the maps $\bullet : M \times M \to M, (x, y) \mapsto xy$, and $\iota : M \to M, x \mapsto x^{-1}$ are smooth. We call a space *connected* if it cannot be divided into disjoint open neighbourhoods. A group is called Abelian or commutative when all elements under the group operation commute.

We apply a theorem from Lie theory, which states that each connected Abelian Lie group $(M, \bullet)$ of dimension $\dim M = n$ is isomorphic to a product

space $\mathbb{R}^p \times \mathbb{T}^q$ with $p + q = n$, for a proof we refer to [23, p. 116]. The q-torus $\mathbb{T}^q$ is a surface of revolution, which means it moves a curve around an axis of rotation. These axes are given by 1-spheres, such that the q-torus is a product space $\mathbb{T}^q = \mathcal{S}^1_1 \times \cdots \times \mathcal{S}^1_q$. The initial decomposition of connected commutative Lie groups can be further simplified into $M \simeq \mathbb{R}^p \times (\mathcal{S}^1_1 \times \cdots \times \mathcal{S}^1_q)$. Recall, that the $(n-1)$-sphere is given by $\mathcal{S}^{n-1} := \{x \in \mathbb{R}^n \mid \|x\|_2 = 1\}$. Thus, $\mathcal{S}^1$ can be embedded into a two-dimensional Euclidean space. Under our assumption, we can derive how many dimensions are at least needed for a suitable embedding. For each 1-sphere we count two dimensions and for each real line accordingly one. Finally, we have to estimate from data how many factor spaces are necessary to represent the topology of the underlying data manifold sufficiently accurately.

3.2 Persistent homology

Algebraic topology provides a computable tool to study not the topology of a set of points itself but an Abelian group which is attached to the topology. The core interest in this discipline lies in *homotopy equivalences* of an equivalence class to which objects belong that are continuously deformable into one another. This term is much broader than a homeomorphism. For two topological spaces X and Y we are looking for a function $h : X \times I \to Y$, which gives the identity at the time $h_0(X) = X$ and for $h_1(X) = X$ a mapping into the other topological space. If this mapping is continuous with respect to its arguments, it is called *homotopy*. Let us consider two functions $f, g : I \to X$ so that $f(1) = g(0)$. Then there is a *composition* of product paths $f \cdot g$, which pass first through f and then through g and which is defined as $f \cdot g(s) = f(2s)$ for $0 \leq s \leq 1/2$ and $g(2s-1)$, for $1/2 \leq s \leq 1$. We first run f at double speed up to $1/2$ and from $1/2$ to 1 we run the function g at double speed. In addition, suppose that we have a family of functions $f : I \to X$ that have the same start and end point $f(0) = f(1) = x_0 \in X$. These functions intuitively form a *loop*. The set of all homotopy classes $[f]$ of loops $f : I \to X$ at base point x_0 is noted as $\pi_1(X, x_0)$ and is called first homotopy group or fundamental group. The group operation is the product $[f][g] = [f \cdot g]$ of equivalence classes. If we do not look at the interval I, but at mappings considering the unit cube I^n, we obtain the n-th homotopy group $\pi_n(X, x_0)$ by analogy. With the help of the homotopy groups we can study the connected components of a topological space for the 0th group, the loops for the 1th group, the cavities for the 2th and so forth. Since homotopy groups are very difficult to compute, we resort to an algebraic variant, the *homology groups*.

Simplices In data analysis, we do not study the topological spaces themselves, but points that we assume are located on or near this space. However, all closed surfaces can be triangulated, i.e. completely covered with simplices. For a proof and illustrations we refer to [14,22, p. 102]. The concept of the triangle is too specific, so the n-*simplices* are defined as a generalization. They are the smallest convex set in Euclidean space with $(n+1)$-points $\nu_0, \cdots, \nu_n$, having no solutions for any system of linear equations in n-variables. Thus, we say they lie in *general*

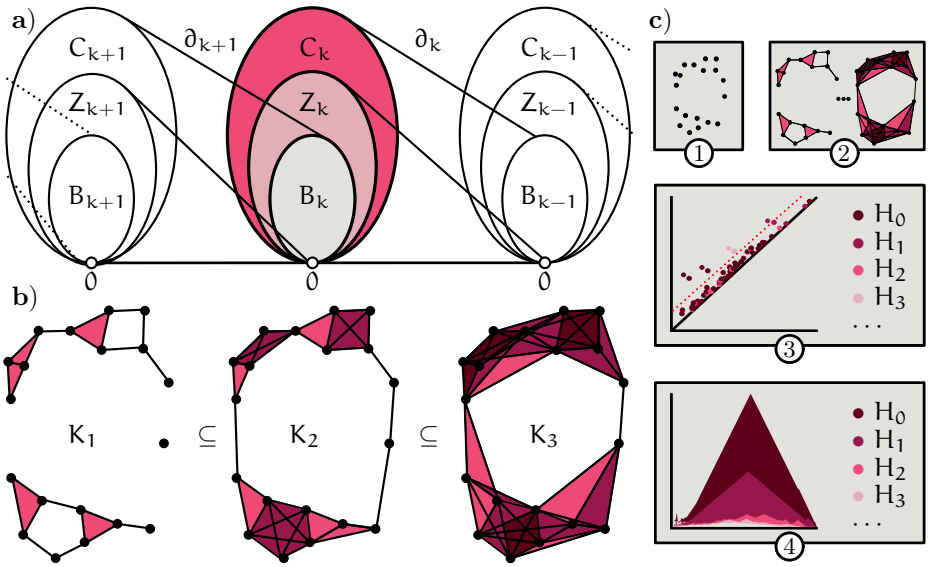


Fig. 1. a) Illustration of the chain complex following Zomorodian et al. [27]. **b**) Sublevel sets of a simplicial complex connected by inclusion. **c**) Pipeline for persistent homology: **c**)1. Loading point sets. **c**)2. Computation of a filtration. **c**)3. Computation of persistent homology with confidence band. **c**)4. Computation of persistence landscapes.

position with respect to $\mathbb{R}^n$, because they do not lie on any hyperplane with dimension less than n [14, p. 103]. We define the n-simplex as follows:

$$[\sigma] = \left\{ [v_0, \cdots, v_n] \in \mathbb{R}^{n+1} \;\middle|\; \sum_{i=0}^{n} v_i = 1 \text{ and } v_i \geq 0 \text{ for all } i \right\}. \tag{1}$$

Removing a vertex from $[\sigma]$ results in a $(n-1)$-simplex called *face* of $[\sigma]$. A 0-simplex is a point, a 1-simplex is a path between two 0-simplices, a 2-simplex is an area enclosed by three 1-simplices and so on. So they follow the intuition and generalize triangles to polyhedra, including points by definition.

Simplicial complexes A *simplicial complex* K is a set of simplices, such that any face of a simplex of K is a simplex of K and the intersection of any two simplices of K is either empty or a common face of both [3, p. 11]. Further, we require every $(k+1)$-element, with $k < n$ to be a k-simplex of $[\sigma] \subset K$. This definition means that the simplicial complex is also a topological space. The properties described ensure that simplices are added to this set in a certain way, namely so that they are only *glued* edge to edge, point to point, and surface to surface. Thus, every n-simplex has $n+1$ distinct vertices and no other n-simplex has the same set of vertices. This is a unique combinatorial description of vertices

together with a collection of sets $\{[\sigma]_1, \cdots, [\sigma]_m\}$ of k-simplices, which are $k+1$ element subsets of K [14, p. 107]. This set system can be realized geometrically with a homeomorphism, by a corresponding mapping into an Euclidean space.

Associated Abelian groups Algebraically, the simplices can be realized as a system of linear combinations as $\sum_i \lambda_i[\sigma]_i$ following Eq. 1. Together with addition as an operation, they form a group structure called kth chain group $\langle C_k, + \rangle$ for the corresponding dimension of the k-simplices used. The group is commutative or Abelian. The used coefficients λ_i induce an orientation, intuitively a direction of the edges is given here by negative and positive signs, so that $[\sigma] = -[\tau]$, iff $[\sigma] = [\tau]$ and $[\sigma]$ and $[\tau]$ have different orientations. The objects of the chain group are called k-chains, for $[\sigma] \in C_k$. The groups are connected by a homomorphism, a mapping which respects the algebraic structure, the boundary operator: $\partial_k : C_k \to C_{k-1}$, $\partial_k[\sigma] \mapsto \sum_i (-1)^i [\nu_0, \nu_1, \cdots, \hat{\nu}_i, \cdots, \nu_n]$. The ith face is omitted, alternatingly. Using the boundary operator we yield the following sequence of Abelian groups, which is a chain complex cf. Fig. 1 a):

$$0 \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} C_{k-1} \xrightarrow{\partial_{k-1}} \cdots \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0. \tag{2}$$

For the calculation of homology groups we have to choose a field from which to obtain the coefficients λ_i. Since, according to the Universal Coefficient Theorem, all homology groups are completely determined by homology groups with integral coefficients, we choose coefficients in $\mathbb{Z}$, see the proof in [13].

The kth chain groups contain two more Abelian subgroups that behave normal to it. First, the so-called cycle groups $\langle Z_k, + \rangle$, which are defined as $Z_k := \ker \partial_k = \{[\sigma] \in C_k \mid \partial_k[\sigma] = \emptyset\}$. This follows the intuition of all elements that form a loop, i.e. have the same start and end point. Some of these loops have the peculiarity of being a boundary of a subcomplex. These elements form the boundary group $\langle B_k, + \rangle$, defined by $B_k := \operatorname{im} \partial_{k+1} = \{[\sigma] \in C_k \mid \exists [\tau] \in C_{k+1} : [\sigma] = \partial_{k+1}[\tau]\}$. Now we define the homology groups as quotient of groups. Homology – analogous to homotopy theory – gives information about connected components, loops and higher dimensional holes in the simplicial complex:

$$H_k(K) := \frac{\ker \partial_k C_k(K)}{\operatorname{im} \partial_{k+1} C_{k+1}(K)} = \frac{Z_k(K)}{B_k(K)}. \tag{3}$$

Homological persistence Examining the homology groups of a set of points gives little information about the structure of a dataset. Instead, we are interested in a parametrization of the simplicial complex as geometric realization in which the homology groups appear and disappear again. For this purpose we consider all possible subcomplexes that form a *filtration* over the point set X. We denote $K^\epsilon := K^\epsilon(X)$. Depending on how we vary the parameter ϵ_i of the chosen simplicial complex, the following sequence is generated, connected by inclusion and starting with the empty set cf. Fig. 1 b):

$$\emptyset = K^{\epsilon_0} \subset K^{\epsilon_1} \subset \cdots \subset K^{\epsilon_{n+1}} = K, \tag{4}$$

$$K^{\epsilon_{i+1}} = K^{\epsilon_i} \cup \sigma^{\epsilon_{i+1}}, \quad \text{for } i \in \{0, \cdots, n-1\}. \tag{5}$$

The filtration has a discrete realization with a fixed $\epsilon = \epsilon_{i+1} - \epsilon_i = \min(x, y)$ for all $x, y \in X$. Through filtration we are able to investigate the homology groups during each step of the parameterization. We record when elements from a homology group appear and when they disappear again. Intuitively speaking, we can see when k-dimensional holes appear and disappear in the filtration. We call this process *birth* and *death* of topological features. Recording the *Betti numbers* of the k-th homology group along the filtration, we obtain the k-*dimensional persistence diagram*, see Fig. 1 c)3. The Betti numbers β_k are defined by rank H_k. We write $H_k^{\epsilon_i}$ as k-th homology group on the simplicial complex K with parametrization ϵ_i. Then $H_k^{\epsilon_i} \to H_k^{\epsilon_{i+1}}$ induces a sequence of homomorphisms on the filtration, for a proof we refer to [10]:

$$0 = H_k^{\epsilon_0} \to H_k^{\epsilon_1} \to \cdots \to H_k^{\epsilon_n} \to H_k^{\epsilon_{n+1}} = 0. \tag{6}$$

The image of each homomorphism consists of all k-dimensional homology classes which are born in K^{ϵ_i} or appear before and die after spanning $K^{\epsilon_{i+1}}$. Tracking the Betti numbers on the filtration results into a multiplicity

$$\mu_k^{\epsilon_i, \epsilon_j} = (\beta_k^{\epsilon_i, \epsilon_{j-1}} - \beta_k^{\epsilon_i, \epsilon_j}) - (\beta_k^{\epsilon_{i-1}, \epsilon_{j-1}} - \beta_k^{\epsilon_{i-1}, \epsilon_j}) \tag{7}$$

for the k-th homology group and index pairs $(\epsilon_i, \epsilon_{j+1}) \in \overline{\mathbb{R}^2} := \mathbb{R}^2 \cup \infty$ with indices $i \leq j$. The Euclidean space is extended, as the very first connected component on the filtration remains connected. Thus, we assign to it infinite persistence, corresponding to the second coordinate ϵ_{j+1}. The first term counts elements born in $K^{\epsilon_{j-1}}$ and which vanish entering K^{ϵ_j}, while the second term counts the representatives of homology classes before K^{ϵ_j} and which vanish at K^{ϵ_j}. The k-th persistence diagram is then defined as

$$\mathrm{PH}_k(X) := \left\{ (\epsilon_i, \epsilon_{j+1}) \in \overline{\mathbb{R}^2} \;\middle|\; \mu_k^{\epsilon_i, \epsilon_{j+1}} = 1 \text{ for all } i, j \in I \right\}. \tag{8}$$

Persistence landscapes Persistence landscapes give a statistical summary of the topology of a set of points embedded in a given topological manifold [4,5]. Looking at the points $(\epsilon_i, \epsilon_{j+1}) \in \overline{\mathbb{R}^2}$ on the k-th persistence diagram $\mathrm{PH}_k(X)$, one associates a piecewise linear function $\lambda_{\epsilon_{j+1}}^{\epsilon_i} : \mathbb{R} \to [0, \infty)$ with those points:

$$\text{If } x \notin (\epsilon_i, \epsilon_{j+1}), \quad \lambda_{\epsilon_{j+1}}^{\epsilon_i}(x) = 0, \tag{9}$$

$$\text{if } x \in \left(\epsilon_i, (\epsilon_i + \epsilon_{j+1})/2\right], \quad \lambda_{\epsilon_{j+1}}^{\epsilon_i}(x) = x - \epsilon_i \text{ and} \tag{10}$$

$$\text{if } x \in \left((\epsilon_i + \epsilon_{j+1})/2, \epsilon_{j+1}\right), \quad \lambda_{\epsilon_{j+1}}^{\epsilon_i}(x) = \epsilon_{j+1} - x. \tag{11}$$

The summaries for the general persistence diagram are the disjoint union of the k-th persistence diagrams $\mathrm{PH}(X) := \coprod_{i=0}^{k} \mathrm{PH}_i(X)$. A persistence landscape $\mathrm{PL}(X)$, contains the birth-death pairs $(\epsilon_i, \epsilon_{j+1})$, for an $i, j \in \{1, \cdots, n\}$ and is the function sequence $\Lambda_k : \mathbb{R} \to [0, \infty)$ for a $k \in \mathbb{N}$, where $\Lambda_k(x)$ denotes the k-th greatest value of $\lambda_{\epsilon_{j+1}}^{\epsilon_i}(x)$, see Fig. 1 c)4. Thus, $\Lambda_k(x) = 0$ if $k > n$. $\mathrm{PL}(X)$ lies in a completely normed vector space, suited for statistical computations.

4 Neural nets

Neural nets are a composition of *affine transformations* with a non-linear activation function. This transformation obtains collinearity. It also preserves parallelism and partial relationships. The projection onto the $(l+1)$-th layer of such a network can be formalized as $\mathbf{x}^{(l+1)} = f(\mathbf{W}^{(l+1)} \cdot \mathbf{x}^{(l)} + \mathbf{b}^{(l+1)})$. The composition of multiple such maps is a deep neural net.

The linear map $\mathbf{x} \mapsto \mathbf{W}\mathbf{x}$ and the statistical distortion term $\mathbf{x} \mapsto \mathbf{x} + \mathbf{b}$ are determined by stochastic gradient descent. Note, that the linear transformation $\mathbf{x} \mapsto \mathbf{W} \cdot \mathbf{x}$ can be interpreted as the product of a matrix $\mathbf{W}_{ij} = \delta_{ij}\mathbf{W}_i$ with the input vector $\mathbf{x}$ using the Kronecker-δ, while $(\cdot)$ denotes elementwise multiplication. Commonly used are functions from the exponential family. As a result a neural net is a map $\varphi^{(l)} : \mathbf{x}^{(l)}(M) \to (\varphi^{(l)} \circ \mathbf{x}^{(l)})(M)$. Learning by back propagation can in this way be considered as a change of coordinate charts of a smooth manifold. For a detailed formulation of back propagation as a shift on the tangent bundle of this manifold we refer to Hauser et al. [15].

In practice neural nets are used with a different number of perceptrons per layer. However, a layer can represent the manifold of the previous one. Manifolds can always be immersed and submersed as long as the rank of the Jacobian of the respective map does not change. Thus, the dimension of the data manifold is equal to the width of the smallest possible layer of a neural net [15].

5 Counting Betti numbers

Based on our assumption that data lies on a connected Abelian Lie group, we can decompose the data space into a product of realaxes and tori. In the following we refer to this Lie group as $G \simeq \mathbb{R}^p \times \mathbb{T}^q$. By this decomposition there is also an isomorphism of the homology groups of G with those of $\mathbb{R}^p \times \mathbb{T}^q$, so that $H_k(G) \simeq H_k(\mathbb{R}^p \times \mathcal{S}^1_1 \times \cdots \times \mathcal{S}^1_q)$ with $p + q = \dim G$. We demand from a neural net that these homology groups can all be approximated, i.e. that all topological structure of the input space can be represented sufficiently well.

One might wonder to what extent the homology groups of a simplicial complex can be used to estimate the homology groups of the manifolds underlying the data. In this section we denote the simplicial homology groups as $H^{\Delta}_{\bullet}(X)$. The $\bullet$ is a proxy for all possible indices. Through the construction of continuous mappings, so-called singular simplices, from a simplex into a topological space, the singular homology groups $H_{\bullet}(X)$ can be constructed. We refer to Hatcher [14, p. 102] for a detailed introduction with proofs. This allows to assign Abelian groups to any topological space. If one constructs smooth mappings instead of continuous ones, one gets smooth singular homology groups $H^{\infty}_{\bullet}(X)$, see for a proof [20, pp. 473]. Finally a chain complex can be formed on a smooth manifold over the p-forms on X and from this the so-called De Rham cohomology can be defined $H^{\bullet}_{dR}(X)$, for a proof we refer to [20, pp. 440]. In summary, we have homology theories for simplicial complexes, topological spaces, a smooth variant on topological spaces and on smooth manifolds. We consider simplicial

homology to be determined on the basis of a simplicial complex, which is merely a rough approximation of the triangulation of the underlying topological space. However, we can draw conclusions about a possible smooth manifold, since the discussed homology groups all have isomorphisms to each other known as the De Rham Theorem, see [14,20, pp. 106, pp. 467]. This legitimizes our approach.

Using Künneth's theorem, we know that the k-th singular homology group of a topological product space $X \times Y$ is isomorphic to the direct sum of the tensor product of its kth homology groups, for a proof we refer to [14, p. 268]:

$$H_k(X \times Y) \simeq \bigoplus_{i+j=k} H_i(X) \otimes H_j(Y). \tag{12}$$

This applies to all fields, since modules over a field are always free. If we apply the theorem to the decomposition, we get for the space we are looking for

$$H_k(\mathcal{S}^1_1 \times \cdots \times \mathcal{S}^1_q \times \mathbb{R}^p) \simeq \tag{13}$$

$$\bigoplus_{i_1+\cdots+i_r=k} H_{i_1}(\mathcal{S}^1_1) \otimes \cdots \otimes H_{i_r}(\mathcal{S}^1_q) \otimes H_{i_r}(\mathbb{R}^p). \tag{14}$$

By the isomorphism of simplicial, singular, smooth singular and De Rham cohomology, we use the theorem in accordance to our assumption. Thus, we count the representatives of homology classes in the persistence landscapes to derive the dimension of the sought manifold. The homology groups of the 1-sphere are

$$H_0(\mathcal{S}^1) = H_1(\mathcal{S}^1) = \mathbb{Z}, \tag{15}$$

$$H_i(\mathcal{S}^1) = 0, \text{ for all } i \geq 2. \tag{16}$$

Note, in Eq. 13 terms remain only for indices $i_j \in \{0, 1\}$. Thus, we get

$$H_0(\mathbb{R}^p) \simeq \mathbb{Z}^p, \tag{17}$$

$$H_k(\mathbb{T}^q) \simeq H_k(\mathcal{S}^1_1 \times \cdots \times \mathcal{S}^1_q) \simeq \mathbb{Z}^{\binom{q}{k}}, \tag{18}$$

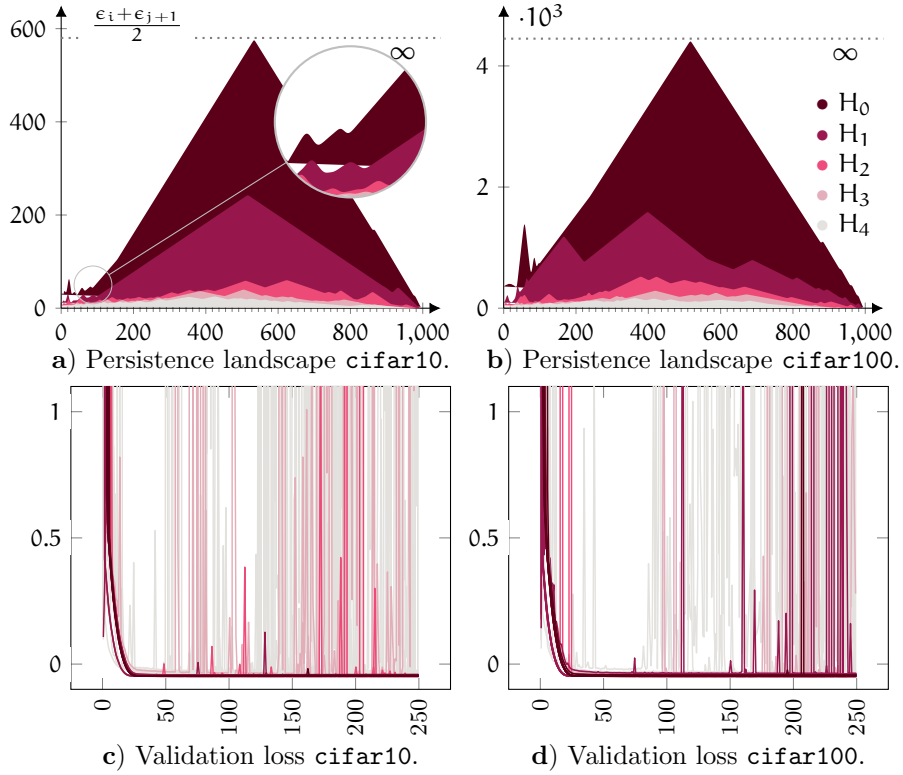
where p indicates the number of connected components. The number of local maxima for a homology group in the persistence landscape is the cardinality of the set of points $(\epsilon_i, \epsilon_{j+1})$ with derivatives $d\lambda^{\epsilon_i}_{\epsilon_{j+1}}(x) = 0$ and $d^2\lambda^{\epsilon_i}_{\epsilon_{j+1}}(x) < 0$, see Fig. 2 b) c). Also, it is the solution for the binomial coefficient:

$$\binom{q}{k} = \frac{q}{1} \cdot \frac{q-1}{2} \cdots \frac{q-(k-1)}{k} = \prod_{i=1}^{k} \frac{q+1-i}{i}. \tag{19}$$

We count in $PL(X)$ the elements of the 0th homology group. The amount of elements from higher homology groups correspond to q (Eq. 18) and are computed using Monte-Carlo methods for an (approximate) integer solution, see Tab. 1.

6 Experimental setting

We train an auto-encoder using as input heavily noisy images and map them to their noiseless original. We perturb each input vector $\mathbf{x}^{(0)}$ with Gaussian noise



a) Persistence landscape `cifar10`. **b**) Persistence landscape `cifar100`.

c) Validation loss `cifar10`. **d**) Validation loss `cifar100`.

Fig. 2. a) Persistence landscape for `cifar10` and b) for the `cifar100` dataset [19]. Persistence landscapes are computed up to a resolution of 10^3. c) and d) show the `MSE` loss function on the validation dataset using a $7:3$ split.

$\epsilon \sim \mathcal{N}(0, \sigma^2\mathbf{I})$, such that the input is weighted $0.5 \cdot \mathbf{x}^{(0)} + 0.5 \cdot \epsilon$. The embedding dimension is increased for each experiment by 2, i.e. $2, 4, 6, \cdots, 784$. 392 neural nets have been trained $\times 100$. Each line of Fig. 2 c), d) represents one neural architecture as averaged loss function.

Persistence landscapes We use the Delaunay-complex for the filtration, according to Melodia et al. [22]. The maximum α-square is set to e^{12}. The maximal expansion dimension of the simplicial complex is set to 10. The maximal edge length is set to 0.1. The persistent landscapes are smoothened by a Gaussian filter $G(x) = 1/\sqrt{2\pi\sigma^2} \cdot e^{-x^2/2\sigma^2}$ with $\sigma = 2$ for visualization purposes. We implement the persistence diagrams and persistence landscapes using the `GUDHI` library [26]. Persistent homology is computed on a `NVIDIA Quadro P4000`.

Hyperparameters The batch size is set to 128, which should help to better recognize good parametrization. Higher batch size causes explosions of the gra-

Table 1. Counts of the representatives per homology group from PL(X).

Data \ Features	Homology groups					≈ embedding dimension					
	H_0	H_1	H_2	H_3	H_4	p	$q\|H_1$	$q\|H_2$	$q\|H_3$	$q\|H_4$	$\dim U$
cifar10	12	16	40	59	50	12	16	$9_{\pm 4}$	$8_{\pm 3}$	$7_{\pm 15}$	$92_{\pm 44}$
cifar100	13	18	34	46	48	13	18	$9_{\pm 2}$	$8_{\pm 10}$	$7_{\pm 13}$	$97_{\pm 50}$

dient, which we visualize in Fig. 2 c), d). For optimization we use `ADAM` [18] and a learning rate of 10^{-4}. The dataset is randomly shuffled. We implement in `Python` 3.7. We use `Tensorflow` as backend [1] with `Keras` as wrapper [6]. The neural training is conducted on `Intel Core i7 9700K` processors.

Neural net We use $L(\mathbf{x}^{(0)}, \mathbf{y}) = 1/n \sum_{i=1}^{n} (\mathbf{x}_i^{(0)} - \mathbf{y}_i)^2$ as loss function. The sigmoid activation function $\sigma(\mathbf{x}^{(l)}) = 1/1 + e^{-\mathbf{x}^{(l)}}$ is applied throughout the net to yield the structure of a smooth manifold. Dense neural nets are used with bias term. The datasets are flattened from $(32, 32, 3)$ into (3072). The first layer is applied to the flattened dataset, which is also adapted by back propagation, i.e. $\varphi^{(0)}(x^{(0)}) = \sigma(\sum_{i=1}^{n} \mathbf{W}_{ji}^{(1)} \mathbf{x}_i^{(0)} + \mathbf{b}_i^{(1)})$. It transfers the data to the desired embedding dimension. The subsequent layers are implemented as residual invertible layers following Dinh et al. [9]. We use invertible architectures to avoid degenerated cases and to stay on the same differentiable structure during learning within the layers to be tested. A total of 5 of these hidden layers are used, all with the same embedding dimension. This diagram describes our architecture:

$$x^{(0)}(M) \xrightarrow{\varphi^{(0)}} x^{(1)}(U) \xleftarrow{\varphi^{(1)}} x^{(2)}(U) \xleftarrow{\varphi^{(L-1)}} \cdots \longrightarrow x^{(L-1)}(U) \xleftarrow{\varphi^{(L)}} x^{(L)}(M).$$

Results In Fig. 2 c) d) we colored the neural nets according to the chosen embedding, such that the lines show the following dimensions ● = $[2, \cdots, 148]$, ● = $[150, \cdots, 198]$, ● = $[200, \cdots, 292]$ and ● = $[\dim U \approx 294, \cdots, 784]$. The embedding dimension can be taken from Tab. 1, due to the trivial invertibility of our neural nets this integer must be doubled, see [9]. All models below our dimensional treshold $\dim U$ show drastic explosions of the gradient, see Fig. 2 c) d). The models above the treshold remain stable. A single net still shows an explosive gradient colored ● in Fig. 2 d). The net showing the deflection of gradient is very close to our threshold, identified as 292 dimensional embedding. The closer we come to the treshold, the sparser the fluctuations of loss. We justified that the minimal embedding dimension does not lie below our treshold.

7 Conclusion

Based on the theory of Lie groups and persistent homology a method for neural nets has been developed to parameterize them using the assumption that the

data lies on or near by some connected commutative Lie group. This forms an approximate solution for a special case of learning problems, allowing to restrict the domain to a commutative group beeing a connected smooth manifold.

Applying Künneth's theorem, the homology groups of the topological factor spaces could be connected with the ones of their product space. Using persistence landscapes the elements originating from some homology groups on the filtration were estimated. With numerical experiments we predicted near ideal embedding dimensions and could confirm, according to Tab. 1, that such a neural embedding delivers a loss function on the validation data set with the smallest fluctuations and the best reconstruction results. We pose following open research questions:

- The neural layers do not have the explicit structure of an Abelian Lie group. How can spherical CNNs [7] be used to always represent a product space $\mathbb{R}^p \times \mathbb{T}^q$ from layer to layer in order to operate on the proposed Lie group?
- This approach can be applied to any kind of manifold, as far as its minimal representation is known. Using decompositions, one may generalize this result. What decomposition allows to neglect the connectedness?

References

1. Abadi, M., et al.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015), https://www.tensorflow.org/
2. Bartlett, P., et al.: Nearly-tight vc-dimension and pseudodimension bounds for piecewise linear neural networks. Journal of Machine Learning Research **20**, 63:1–63:17 (2019)
3. Boissonnat, J.D., et al.: Geometric and Topological Inference. Cambridge University Press (2018)
4. Bubenik, P.: Statistical topological data analysis using persistence landscapes. Journal of Machine Learning Research **16**, 77–102 (2015)
5. Bubenik, P., Dlotko, P.: A persistence landscapes toolbox for topological statistics. Journal of Symbolic Computation **78**, 91–114 (2017)
6. Chollet, F., et al.: Keras. https://keras.io (2015)
7. Cohen, T.S., et al.: Spherical cnns. In: 6th International Conference on Learning Representations (2018)
8. Cybenko, G.: Approximation by superpositions of a sigmoidal function. MCSS **5**(4), 455 (1992)
9. Dinh, L., et al.: Density estimation using real NVP. In: 5th International Conference on Learning Representations (2017)
10. Edelsbrunner, H., Harer, J.: Persistent homology – a survey. Contemporary mathematics **453**, 257–282 (2008)
11. Esteves, C., et al.: Learning SO(3) equivariant representations with spherical cnns. Int. J. Comput. Vis. **128**(3), 588–600 (2020)
12. Futagami, R., Yamada, N., Shibuya, T.: Inferring underlying manifold of data by the use of persistent homology analysis. In: 7th International Workshop on Computational Topology in Image Context. pp. 40–53 (2019)
13. Gruenberg, K.: The universal coefficient theorem in the cohomology of groups. Journal of the London Mathematical Society **1**(1), 239–241 (1968)
14. Hatcher, A.: Algebraic Topology. Cambridge University Press (2002)

15. Hauser, M., Ray, A.: Principles of riemannian geometry in neural networks. In: Adv Neural Inf Process Syst 30. pp. 2807–2816 (2017)
16. Hauser, M., et al.: State-space representations of deep neural networks. Neural Computation **31**(3) (2019)
17. Johnson, J.: Deep, skinny neural networks are not universal approximators. In: 7th International Conference on Learning Representations (2019)
18. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations (2015)
19. Krizhevsky, A.: Learning multiple layers of features from tiny images (2009)
20. Lee, J.: Introduction to smooth manifolds. Springer (2013)
21. Lin, H., Jegelka, S.: Resnet with one-neuron hidden layers is a universal approximator. In: Adv Neural Inf Process Syst 31. pp. 6172–6181 (2018)
22. Melodia, L., Lenz, R.: Persistent homology as stopping-criterion for voronoi interpolation. In: 20th International Workshop on Combinatorial Image Analysis (2020)
23. Onishchik, A., et al.: Lie groups and Lie algebras. Springer (1993)
24. Raghu, M., et al.: On the expressive power of deep neural networks. In: 34th ICML. pp. 2847–2854 (2017)
25. Stone, M.: The generalized weierstrass approximation theorem. Mathematics Magazine **21**(5), 237–254 (1948)
26. The GUDHI Project: GUDHI User and Reference Manual. GUDHI Editorial Board, 3.1.1 edn. (2020), https://gudhi.inria.fr/doc/3.1.1/
27. Zomorodian, A., Carlsson, G.: Computing persistent homology. Discrete Computational Geometry **33**(2), 249–274 (2005)