

# Parametrization of Invertible Resnets up to Homology Equivalence for Abelian Lie Groups

Luciano Melodia<sup>[0000–0002–7584–7287]</sup>  
and Richard Lenz<sup>[0000–0003–1551–4824]</sup>

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

**Abstract.** Neural nets have been used in an elusive number of scientific disciplines. Nevertheless, their parameterization is largely unexplored. Dense nets are the coordinate transformations of a manifold from which the data is sampled. After processing through a layer, the representation of the original manifold may change. This is crucial for the preservation of its topological structure and should therefore be parameterized correctly. We discuss a method to determine the smallest topology preserving layer for an invertible residual net. We consider the data domain as abelian connected matrix group and observe that it is decomposable into  $\mathbb{R}^p \times \mathbb{T}^q$ . Persistent homology allows us to count its  $k$ -th homology groups. Using Künneth’s theorem, we count the  $k$ -th Betti numbers. Since we know the embedding dimension of  $\mathbb{R}^p$  and  $S^1$ , we parameterize the bottleneck layer with the smallest possible matrix group, which can represent a manifold with those homology groups. Resnets guarantee smaller embeddings due to the dimension of their state space representation.

**Keywords:** Embedding Dimension · Parametrization · Persistent Homology · Neural Nets · Manifold Learning.

## 1 Motivation

Neural nets and their underlying principles were studied from the perspective of differential and information geometry [25]. The natural gradient descent turned out to be Fisher optimal [1,2], in its interpretation as a walk on the hypersurface of a neural loss function. Local minima, maxima and saddle points on the hypersurface have always been a problem when training machine learning models. A critical point on this surface corresponding to the global minimum of a smaller model forms a local minimum or saddle point of a larger one [11]. In other words, the critical values of a function represented by a layer of  $(n - 1)$ -perceptrons will result in many critical values of a layer of  $n$ -perceptrons.

The manifold of a dense neural net is in most cases Euclidean, but it is restricted by the loss function or the data manifold to certain matrix groups. A metric is required to describe the geometric properties of such a manifold.

Recently [14], the change of coordinate systems and the learned metric tensor of a Riemann manifold during back propagation of neural nets was formulated. Further, it has been shown that dense neural nets can't approximate all functions arbitrarily precisely [16]. In response to this result, neural nets with residual connections were examined. Resnets 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 an universal function approximator [19]. The depth of the net is a parameter depending on the moduli of continuity of a function to be approximated [19]. Such a residual layer defines 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 [13]. Estimating simplicial homology in connection with Künneth's theorem we found a resnet with an approximate minimal number of perceptrons for a very special learning problem.

We use invertible architectures, to demonstrate our hypothesis on a different lie Group from Euclidean space [21]. Our interpretation of a neural net is a continuous transformation, transferring invariant structures of the data in the input space to its deepest layer. Thus, it is capable to reconstruct the structure of the original space. This approach corresponds to the concept of homotopy equivalence [22,26], which means that there is a continuous map  $f : X \rightarrow Y$  and  $g : Y \rightarrow X$ , such that  $f \circ g$  is homotopic to the identity  $\text{Id}_Y$  and  $g \circ f$  to  $\text{Id}_X$ . Thus, the spaces can be deformed continuously into one another. To do so, their homotopy groups must be equal. The fundamental group gives information about connected components which can be collapsed to a point by continuous functions, the first homotopy group about 'holes' and the second about 'cavities' and so forth. But homotopies are difficult to compute [24,34]. Homology theory gives a remedy that guarantees an isomorphism between homotopy groups and homology groups under the conditions of Hurewicz's Theorem [24]. We use theoretical results from Lie theory and persistent homology theory to estimate how many dimensions of embedding are needed to represent the homology groups. Our paper is structured as follows:

- In Sect. 2 we discuss the mathematical assumptions and recall the principles of smooth manifolds and Lie theory. In particular, we recall the decomposition of connected commutative Lie groups into  $\mathbb{R}^p \times \mathbb{T}^q$ .
- In Sect. 3 we introduce to the theory of persistent homology, define persistence diagrams, persistence landscapes and confidence sets, in order to estimate homology groups despite noisy data. We use Künneth's theorem and derive how to count the homology groups of the input space.
- In Sect. 4 we explain the mathematical theory of neural nets and construct invertible deep neural nets to satisfy our assumptions. Resnets, which have a larger state space dimension, offer an architecture with a comparatively small number of parameters to dense nets.
- In Sect. 5 we validate our theory training autoencoders on the cifar10 and cifar100 datasets, isolating the innermost layer such that no residual connections bypass it on different embedding dimensions.

## 2 The Manifold Assumption

The manifold assumption states that a point  $x = \{x_1, \dots, x_d\} \in \mathbb{R}^n$  lies on an embedded manifold  $M$ , with  $\dim M \ll n$ . We specify the concept of smooth manifolds. Then we dive into Lie groups concluding with a central theorem on the isomorphism of abelian connected Lie groups to the product space  $\mathbb{R}^p \times \mathbb{T}^q$ , which turns out useful to estimate the embedding dimension of the manifold.

### 2.1 Background: Smooth Manifolds

Manifolds generalize the concept of space mathematically. They are described by a collection of local coordinate maps  $\varphi : U \rightarrow \varphi(U) \subseteq \mathbb{R}^n$  which are homeomorphic to  $\mathbb{R}^n$  where  $n = \dim M$  and an open neighborhood  $(U_i, \varphi_i)$  in  $M$  which is covered by this map. The union of such maps should cover the entire space, i.e.  $M = \bigcup_{i \in I} U_i$  and is called *atlas*  $\mathcal{A} = \{(U_i, \varphi_i) \mid i \in I\}$  [37]. The functions  $x_1, \dots, x_n : U \rightarrow \mathbb{R}$  are called local coordinates  $\varphi(p) = (x_1(p), \dots, x_n(p))$ .

We require that the coordinate system on  $M$  is smooth. This is due to the condition that two maps of a differentiable structure must always be compatible with each other. Compatibility means, that the maps  $(U, \varphi), (V, \psi)$  are diffeomorphic, such that their composition  $\psi \circ \varphi^{-1} : \varphi(U \cap V) \rightarrow \psi(U \cap V)$  is a diffeomorphism between the two open subsets  $\varphi(U \cap V), \psi(U \cap V) \subseteq \mathbb{R}^n$ , i.e.  $\psi \circ \varphi^{-1}$  and  $\varphi \circ \psi^{-1}$  are in  $\mathcal{C}^\infty(M)$ . Therefore, they are also smooth. If we take a step back and neglect smoothness for  $M$ , then  $M$  is a topological manifold.

If we now equip  $M$  with such an atlas  $\mathcal{A}$  and if this 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  $\mathcal{A}_{max}$  [37]. Thus, it is sufficient to describe smooth manifolds by an atlas that is not maximal.

### 2.2 Background: Lie groups

A Lie group is a  $\mathcal{C}^\infty$ -manifold  $G$  which is equipped with a group structure such that the maps  $\mu : (x, y) \mapsto xy, G \times G \rightarrow G$  and  $\iota : x \mapsto x^{-1}, G \rightarrow G$  are smooth. In our case  $\mu$  corresponds to the matrix multiplication. The neutral element of a group is denoted by  $e = e_G$ . Each abelian commutative Lie group  $G$  of dimension  $\dim G = m$  is isomorphic to a particular product space  $\mathbb{R}^p \times \mathbb{T}^q$ , with  $p + q = m$ , see Appx. A.1. Thus, the group can be decomposed into a product space with a relatively simple topological structure. The  $q$ -torus  $\mathbb{T}^q$  is a surface of revolution. It moves a curve around an axis of rotation. These axes are given by 1-spheres, so the  $q$ -torus is the product space  $\mathbb{T}^q := \mathcal{S}_1^1 \times \dots \times \mathcal{S}_q^1$ . Our initial decomposition of the Lie group into a product space can be further simplified into  $G \simeq \mathbb{R}^p \times (\mathcal{S}_1^1 \times \dots \times \mathcal{S}_q^1)$ . Note, 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. The homology groups and the embedding dimensions of the product spaces, yield by decomposition of connected commutative Lie groups, are therefore known. Any Euclidean space like  $\mathbb{R}^p$  is contractible, i.e. zero-homotopic

or homotopy equivalent to a constant map, thus the homology group  $H_k(\mathbb{R}^p) = 0$  for  $k > 0$  and  $H_k(\mathbb{R}^p) = \mathbb{Z}$  for  $k = 0$ . To illustrate the role of orientation, we calculate the homology groups and illustrate the triangulation of the 1-sphere, both with coefficients in  $\mathbb{R}$  and in  $\mathbb{Z}/2\mathbb{Z}$  in Fig. 1. The latter field neglects orientation. Note, that for this field  $+1 = -1$ .

A tool is needed to estimate homology groups of the manifold on which the data points lie. The method can then be inversely applied to the homology groups to estimate its embedding dimension. We introduce persistent homology with its confidence intervals in order to obtain the best possible estimate of the current state of the art by removing topological noise.

### 3 Persistent Homology

In algebraic topology spaces are studied as the composition of triangles, the latter being used as the fundamental component of a large class of spaces. All closed surfaces can be constructed by identifying – or, less formally, ‘gluing’ – the edges of triangles [34, 20, p. 102]. The  $n$ -dimensional generalization of triangles are  $n$ -simplices. They are the smallest convex set in Euclidean space with  $(n+1)$ -points  $v_0, \dots, v_n$ , which have no solutions for any system of linear equations in  $n$ -variables. Thus, they do not lie on any hyperplane with a dimension less than  $n$  [34, p. 103]. We define the  $n$ -simplex as follows:

$$\Delta^n = \left\{ [v_0, \dots, v_n] \in \mathbb{R}^{n+1} \mid \sum_{i=0}^n v_i = 1 \text{ and } v_i \geq 0 \text{ for all } i \right\}. \quad (1)$$

Removing a vertex from  $\Delta^n$  results in a  $(n-1)$ -simplex called *face* of  $\Delta^n$ . The combinatorial description of  $n$ -simplices is called *simplicial complex*.

#### 3.1 Simplicial Complexes and Homology Theory

Simplicial homology is studied by  $\Delta$ -complexes whose simplices are uniquely determined by their vertices. Thus, every  $n$ -simplex has  $(n+1)$  distinct vertices and no other  $n$ -simplex has the same set of vertices. A *simplicial complex*  $K$  is a collection 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 [30, p. 11]. This is a unique combinatorial description of vertices together with a collection of sets  $\{\Delta_1^k, \dots, \Delta_m^k\}$  of  $k$ -simplices, which are  $(k+1)$  element subsets of  $K$  [34, p. 107]. Further, we require every  $(k+1)$ -element, with  $k < n$  to be a  $k$ -simplex of  $\Delta^k \subset K$ . Simplicial complexes have a natural order of simplices, which is illustrated in Fig. 1. It plays a central role in their geometric realization [30, p. 14]. If we choose a total order for all corners which are located in any simplex of  $K$ , we get an order for each single  $k$ -simplex.

A *singular  $k$ -simplex* is defined as a continuous map  $\sigma : \Delta^k \rightarrow X$  into a topological space  $X$ . Intuitively, every simplicial complex  $K$  has some connected components. These form a vector space  $C_k(X)$ , which is called  *$k$ -th chain group*,

with coefficients in  $\mathbb{Z}$  and elements called *k-chains*. The *k-chains* are finite sums  $\sum_{i=0}^k z_i \sigma_i$  with  $z_i \in \mathbb{Z}$  [34, p. 108]. Having connected components defined as a group implies also the definition of a boundary. As the boundaries of a *k-simplex* are also  $(k-1)$ -simplices, we can remove all other simplices to project into the lower chain group by the *k-th* boundary operator  $\partial_k : C_k(X) \rightarrow C_{k-1}(X)$ :

$$\partial_k(\sigma) = \sum_{i=0}^k (-1)^i \sigma \Big| [v_0, \dots, \hat{v}_i, \dots, v_k]. \quad (2)$$

Observe that the composition of boundary maps is zero, as a boundary has no boundary by itself, i.e.  $\partial_i \partial_{i+1} = 0$  [34, pp. 108–109]. Thus,  $\ker \partial_i$  corresponds to the generating elements of the *i-th* boundary group and  $\text{im } \partial_{i+1}$  for the *i-th* chain group. The singular homology group is the quotient space

$$H_k(X) = \frac{\ker(\partial_k : C_k(X) \rightarrow C_{k-1}(X))}{\text{im}(\partial_{k+1} : C_{k+1}(X) \rightarrow C_k(X))}. \quad (3)$$

The *k-th* singular homology group is isomorphic to the *k-th* simplicial homology group for a triangulable metric space, i.e.  $H_k(X) \simeq H_k^\Delta(X)$  for all  $k \in \mathbb{Z}$ , see Appx. B.3, thus we can transfer results from one homology theory to the other.

Examining the homology groups of a set of points gives us little information about the structure of our dataset. However, since the homology groups of a simplicial complex are equal to the direct sum of the ones of all disjoint subcomplexes, see Appx. B.2, we examine them more closely. In particular, 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 cloud  $X$ . Depending on how we vary the parameter  $\omega_i$  of the chosen simplicial complex, the following sequence is generated, starting with the empty set:

$$\emptyset = K^{\omega_0}(X) \subset K^{\omega_1}(X) \subset \dots \subset K^{\omega_{n+1}}(X) = K(X), \quad (4)$$

$$K^{\omega_{i+1}}(X) = K^{\omega_i}(X) \cup \sigma^{\omega_{i+1}}, \quad \text{for } i \in \{0, \dots, n-1\}. \quad (5)$$

The filtration has a discrete realization with a fixed  $\epsilon = \omega_{i+1} - \omega_i = \min(x, y)$  for all  $x, y \in X$ . We chose a sufficiently small  $\epsilon$ , in order to capture the birth and death of critical values in the sublevel sets of the simplicial complex upon the data, to be able to map as much structure as possible into the persistence diagram – which clearly depends on the metric.

### 3.2 Persistence Diagrams

Recording the *Betti numbers* of the *k-th* homology group along the filtration, we obtain the *k-dimensional persistence diagram*. The Betti numbers  $\beta_k$  are defined by  $\text{rank}(H_k(X))$ . We write  $H_k^{\omega_i} = H_k(K^{\omega_i}(X))$  as *k-th* homology group on the simplicial complex  $K$  with parametrization  $\omega_i$ . Then  $H_k^{\omega_i} \rightarrow H_k^{\omega_{i+1}}$  induces a sequence of homomorphisms on the filtration, see Appx. B [20,9]:

$$0 = H_k^{\omega_0} \rightarrow H_k^{\omega_1} \rightarrow \dots \rightarrow H_k^{\omega_n} \rightarrow H_k^{\omega_{n+1}} = 0. \quad (6)$$



**Fig. 1.** Two triangulations of  $S^1$ . We compute  $H_0(S^1) = \ker(\partial_0)/\text{im}(\partial_1)$  for (a) and (b), see Sect. 3.1. For (a):  $\ker(\partial_0) = \text{span}(\{1\}, \{2\}, \{3\})$  and  $\text{im}(\partial_1) = \text{span}(\{\{2\} - \{1\}\}, \{\{3\} - \{2\}\})$  as linearly independent elements. Observe that  $\{1\} - \{3\} = \{2\} - \{1\} - (\{3\} - \{2\})$  holds. Thus  $\ker(\partial_0) \simeq (\mathbb{Z}/2\mathbb{Z})^3$ ,  $\text{im}(\partial_1) = (\mathbb{Z}/2\mathbb{Z})^2$  and  $H_0(S^1; \mathbb{Z}/2\mathbb{Z}) = \mathbb{Z}/2\mathbb{Z}$ . For (b):  $\ker(\partial_0) = \text{span}(\{1\}, \{2\}, \{3\})$ , but due to real coefficients the linear dependency is this time given by  $\{2 - 1\} = \{3 - 2\}$ . Thus  $\ker(\partial_0) \simeq \mathbb{R}^3$  and  $\text{im}(\partial_1) = \mathbb{R}^2$ , which yields the homology group  $H_0(S^1; \mathbb{R}) = \mathbb{R}$ .

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

$$\mu_k^{\omega_i, \omega_j} := (\beta_k^{\omega_i, \omega_{j-1}} - \beta_k^{\omega_i, \omega_j}) - (\beta_k^{\omega_{i-1}, \omega_{j-1}} - \beta_k^{\omega_{i-1}, \omega_j}) \quad (7)$$

for the  $k$ -th homology group and index pairs  $(\omega_i, \omega_{j+1}) \in \overline{\mathbb{R}^2} := \mathbb{R}^2 \cup \infty$  with indices  $i \leq j$  [9]. The Euclidean space is extended, as the first connected component of our filtration is not merged with any other critical value. We assign the value  $\infty$  to it, corresponding to the second coordinate  $w_{j+1}$  in the persistence diagram. The first term counts elements born in  $K^{\omega_{j-1}}(X)$  and vanished entering  $K^{\omega_j}(X)$ , while the second term counts the representatives of homology classes before  $K^{\omega_j}(X)$  and vanishing at  $K^{\omega_j}(X)$ . The index pairs denote the *birth* and *death* of a homology class. The  $k$ -th persistence diagram is thus

$$\mathfrak{P}_k(K(X)) := \left\{ (\omega_i, \omega_{j+1}) \in \overline{\mathbb{R}^2} \mid \mu_k^{\omega_i, \omega_{j+1}} = 1 \text{ for all } i, j \in I \right\}. \quad (8)$$

### 3.3 Confidence Sets

We are interested in the homology groups of  $M$  and not in those of  $K(X)$ . Nevertheless, the point set  $X$  is drawn from a distribution, that is either on or near  $M \subset \mathbb{R}^d$ . We need to decide whether a persistent homology class comes from this distribution or not. For this purpose, confidence intervals for persistence diagrams have been defined [5, 10]. A confidence interval starts at the off-diagonal of a persistence diagram and ends parallel to it. It determines which topological features are persistent enough, i.e. persist in the filtration long enough to be recognized as such. Otherwise, they are treated as topological noise and removed.

For two compact sets  $X, Y \subset \mathbb{R}^n$ , the persistent homology on their filtration is stable with respect to the Hausdorff distance [10], which is defined as

$$d_H(X, Y) := \max \left\{ \sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y) \right\}. \quad (9)$$

Furthermore, persistent homology is stable with respect to the bottleneck distance – an extremal case of the Wasserstein distance – for some  $p \rightarrow \infty$ :

$$d_B(X, Y) := \inf_{\varphi: X \rightarrow Y} \sup_{x \in X} \|x - \varphi(x)\|_p, \quad (10)$$

with  $\varphi$  being within the set of all bijections from  $X$  into  $Y$ , yielding

$$d_B(\mathfrak{P}(K(X)), \mathfrak{P}(K(Y))) \leq c \cdot d_H(X, Y). \quad (11)$$

Let  $\mathbb{P}$  be a distribution of a finite set of points  $X$  and  $x \in X$  a sample drawn from this distribution. Then we can define a confidence interval with respect to the Hausdorff distance as  $\mathbb{P}(d_H(x, X) \leq \mu_\alpha) \geq 1 - \alpha$  [10]. A persistence confidence set  $\mathcal{C}(K(X))$  is a subset of all persistence diagrams which have a certain distance from the diagonal. We look for all homology classes persisting a minimum ‘time’:

$$\mathcal{C}(K(X)) := \left\{ \mathfrak{P}(K(X)) \mid c \cdot d_B(\text{Diag}(\mathfrak{P}(K(X))), \mathfrak{P}(K(X))) \leq \mu_\alpha \right\}. \quad (12)$$

Recall, that no points can appear under the diagonal line. Therefore, we measure the distance to the diagonal. For  $d_B(x, y)$  imagine a box around each point with a certain radius  $\mu$ . If the diagonal of the persistence diagram passes through the box of a point, the point is considered as noise and removed.

### 3.4 Persistence Landscapes

The summaries of topologically persistent features – with the purpose of computing statistics in a Hilbert space – was initiated by Bubenik [5]. Bubenik and Dlotko first provided efficient algorithms to compute persistence landscapes [6]. Persistence landscapes give a summary of the topology of a set of points embedded in a given topological manifold. Looking at the points  $(\omega_i, \omega_{j+1}) \in \overline{\mathbb{R}^2}$  on the  $k$ -th persistence diagram  $\mathfrak{P}_k(K(X))$ , one associates a piecewise linear function  $f_{(\omega_i, \omega_{j+1})} : \mathbb{R} \rightarrow [0, \infty)$  with them, which is defined as

$$f_{(\omega_i, \omega_{j+1})}(x) = \begin{cases} 0, & \text{if } x \notin (\omega_i, \omega_{j+1}), \\ x - \omega_i, & \text{if } x \in \left( \omega_i, \frac{\omega_i + \omega_{j+1}}{2} \right], \\ \omega_{j+1} - x & \text{if } x \in \left( \frac{\omega_i + \omega_{j+1}}{2}, \omega_{j+1} \right). \end{cases}$$

We describe the summaries for the general persistence diagram, being the disjoint union of the  $k$ -th persistence diagrams  $\mathfrak{P}(K(X)) := \coprod_{i=0}^k \mathfrak{P}_i(K(X))$ . A

persistence landscape  $\mathfrak{L}(\mathfrak{P}(K(X)))$ , contains the birth-death pairs  $(\omega_i, \omega_{j+1})$ , for an  $i, j \in \{1, \dots, n\}$  and is the function sequence  $\lambda_k : \mathbb{R} \rightarrow [0, \infty)$  for a  $k \in \mathbb{N}$ , where  $\lambda_k(x)$  denotes the  $k$ -th greatest value of  $f_{(\omega_i, \omega_{j+1})}(x)$ . Thus,  $\lambda_k(x) = 0$  if  $k > n$ . In contrast to standard topological summaries, the persistence landscape lies in a completely normed vector space, suited for statistical computations.

### 3.5 Counting Betti Numbers

In the following, the homology groups are computed over  $\mathbb{Z}$ . Based on our assumption to operate on a connected commutative Lie group, there is an isomorphism for the homology groups of the data space  $G$ , so that  $H_k(G) \simeq H_k(\mathbb{R}^p \times \mathcal{S}_1^1 \times \dots \times \mathcal{S}_q^1)$  with  $p + q = \dim G$ , see A.3. The homology groups of this product space are equal to the homology groups of the sought manifold. The neural net should be able to preserve all structures within this space. Since the input domain is a set of points to be described by one or more compact manifolds, all homology classes should be preserved in our learning task. Künneth's theorem states that the  $k$ -th singular homology group of a topological product space  $X \times Y$  is isomorphic to the product of the respective  $k$ -th homology groups of  $X$  and  $Y$ , see [34, p. 268] and Appx. A.1:

$$H_k(X \times Y) \simeq \bigoplus_{i+j=k} H_i(X) \otimes H_j(Y). \quad (13)$$

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

$$H_k(\mathcal{S}_1^1 \times \dots \times \mathcal{S}_q^1 \times \mathbb{R}^p) \simeq \quad (14)$$

$$\simeq \bigoplus_{i_1 + \dots + i_r = k} H_{i_1}(\mathcal{S}_1^1) \otimes \dots \otimes H_{i_r}(\mathcal{S}_q^1) \otimes H_{i_r}(\mathbb{R}^p). \quad (15)$$

We need to know how to count in order to derive the dimension of the sought manifold from the representatives of homology classes within the mean persistence landscapes. Note, in Eq. 14 terms remain only for indices  $i_j \in \{0, 1\}$ . Thus, the homology group  $H_k(\mathbb{T}^q) \simeq H_k(\mathcal{S}_1^1 \times \dots \times \mathcal{S}_q^1) \simeq \mathbb{Z}^{\binom{q}{k}}$  and  $H_0(\mathbb{R}^p) \simeq \mathbb{Z}^p$ , where  $p$  indicates the number of connected components. We denote the amount of representatives from homology classes of the  $k$ -th homology group appearing in the mean persistence landscape  $\bar{\mathfrak{L}}$  by  $\sharp_k^H(\bar{\mathfrak{L}}(\mathfrak{P}(K(X))))$ . Thus, we get the following relation for the representatives of the zeroth homology group:

$$\sharp_0^H(\bar{\mathfrak{L}}(K(X))) \approx p. \quad (16)$$

This is the number of connected components to be represented by the innermost layer of the neural net, corresponding exactly to the number of elements from the zeroth homology group. We only need to consider the homology groups which are represented in the mean persistence landscape. Assuming that our estimate





**Fig. 2.** The architecture of the neural net maps over  $\varphi^{(0)}$  into a subspace of  $M$ , in which the searched manifold  $U$  of the data can be embedded. The maps  $\varphi^{(1)}, \dots, \varphi^{(L-1)}$  are models of bijective, smooth and therefore diffeomorphic functions.

is good, higher homology groups are trivial. The required number of factors of 1-spheres can be determined for the  $k$ -th homology group by

$$\sharp_k^H(\bar{\mathfrak{L}}(\mathfrak{P}(K(X)))) \approx \prod_{i=1}^k \frac{q - k + i}{i}. \quad (17)$$

Note, that we assume to work on a smooth manifold, whose de Rham cohomology induces an isomorphism into the simplicial homology with coefficients in  $\mathbb{R}$ , see Appx. B.3. We can still work with coefficients in  $\mathbb{Z}$  because, according to the Universal Coefficient Theorem, the integral homology groups determine the homology groups over every field [33]. Thus, these results can be applied without restriction to any homology theory, see Appx. B.3.

## 4 Neural Nets

Neural nets are a composition of affine transformations with a non-linear activation function. The projection on the  $(l+1)$ -th layer can be formalized as

$$\varphi^{(l+1)} := \mathbf{x}^{(l+1)} = \sigma(\mathbf{W}^{(l+1)} \cdot \mathbf{x}^{(l)} + \mathbf{b}^{(l+1)}). \quad (18)$$

The linear map  $\mathbf{x} \mapsto \mathbf{W}\mathbf{x}$  and the statistical distortion term  $\mathbf{x} \mapsto \mathbf{x} + \mathbf{b}$  is determined by stochastic gradient descent. Note, that the linear transformation  $\mathbf{x} \mapsto \mathbf{W} \cdot \mathbf{x}$  is interpreted as the product of a matrix  $\mathbf{W}_{ij} = \delta_{ij} \mathbf{W}_i$  with the input vector  $\mathbf{x}$  and the Kronecker delta, while  $(\cdot)$  denotes elementwise multiplication.

$\sigma$  denotes a non-linear activation function. Commonly used are functions from the exponential family. A feedforward net is then a map  $\varphi^{(l)} : \mathbf{x}^{(l)}(M) \rightarrow (\varphi^{(l)} \circ \mathbf{x}^{(l)})(M)$ .  $\mathbf{x}^{(0)} \in M$  are points on a Riemann manifold  $M$ . Learning by back propagation can be considered as a change of coordinate representation or shift on the tangent bundle of the underlying manifold [23].

### 4.1 Deep Learning

Neural nets are used in practice with a different number of perceptrons per layer. However, a layer, from the perspective of a matrix group, can also 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 remains constant. Thus, the dimension of the data manifold is the dimension of the smallest possible bottleneck layer of the neural net  $\dim M = \min_l \dim \mathbf{x}^{(l)}$  [14]. Our contribution provides the theory of how this dimension can be determined for all neural nets that are a composition of invertible layers acting on an abelian Lie group as domain of the data. The concatenation of such maps is called a deep neural net:

$$\mathbf{x}^{(L)}(M) = \left( \varphi^{(l+1)} \circ \varphi^{(l)} \dots \circ \varphi^{(1)} \circ \varphi^{(0)} \circ \mathbf{x}^{(0)} \right) (M). \quad (19)$$

Activation functions restrict to a subgroup of Euclidean space. Continuous homomorphisms between Lie groups are automatically smooth. Thus, a second countable locally Euclidean group has exactly one differentiable structure, turning it into a Lie group. We need to show that  $\mathbf{x}^{(l+1)}$  has the topology of  $G$  and is therefore an embedded closed Lie subgroup  $\mathbf{x}^{(l)} := H \subseteq G$ . Cartan's Theorem of Closed Subgroups states that a subgroup  $H$  of a Lie group  $G$  is an embedded Lie subgroup, if it is a closed subgroup [7]. We construct our neural nets such that they are a subset of  $\text{GL}(n, \mathbb{R})$  and according to the proof in Appx. 2.2, this is a closed subgroup of  $\text{GL}(n, \mathbb{C})$ , which is itself a Lie group. It follows that a layer in the neural net is also a Lie group. Due to invertibility, the subsequent smaller layers are Lie subgroups, as  $\sigma^{(l)}$  is smooth.

## 4.2 Residual Neural Networks

A *resnet* skips some of the layers by adding outputs to a deeper layers. It has one *skip connection* and is a  $\mathcal{C}^1$ -continuous dynamic transformation [15,13]:

$$\mathbf{x}^{(l+1)} = \mathbf{x}^{(l)} + \sigma^{(l)} \left( \mathbf{x}^{(l)} \right) \Delta l. \quad (20)$$

The term  $\Delta l$  in Eq. 20 is inserted as a perturbation term. In the work of Hauser et al. [14], neural nets were classified as  $\mathcal{C}^k$ -architecture, depending on how often a *finite difference operator* is applied to the map  $\mathbf{x} : M \times I \rightarrow \mathbb{R}^d$ . The forward and backward difference operators are defined as  $\delta^+ : \mathbf{x}^{(l)} \mapsto \mathbf{x}^{(l+1)} - \mathbf{x}^{(l)}$  and  $\delta^- : \mathbf{x}^{(l)} \mapsto \mathbf{x}^{(l)} - \mathbf{x}^{(l-1)}$ . In this theory, we can write the difference operators for a general  $n \in \mathbb{N}$  with explicit specification of the perturbation  $\Delta l$  as

$$\delta^+ (\delta^-)^{n-1} \mathbf{x}^{(l)} = \sum_{l'=0}^k \left[ (-1)^{l'} \binom{k}{l'} \mathbf{x}^{(l+1-l')} \right] = \sigma^{(l)} \left( \mathbf{x}^{(l)} \right) \Delta l^k. \quad (21)$$

We use the same notation as in [13], where  $(\delta^-)^{k-1} := \delta^- \delta^- \dots \delta^-$  denotes the application of the difference operator  $(k-1)$ -times and the binomial coefficient  $\binom{k}{l'}$ , with  $l'$  indicating the index for summation over the layers. Using one forward difference operator and the remaining  $(k-1)$  backward difference operators, the next layer  $\mathbf{x}^{(l+1)}$  is a function of the  $k$  previous layers  $\mathbf{x}^{(l)}, \mathbf{x}^{(l-1)}, \dots, \mathbf{x}^{(l-k+1)}$ . To calculate the *state space dimension* of equations of degree  $k$  as in Eq. 21, we

define the state as the difference operators of different transformation order of an input vector  $\mathbf{x}$  in the layer  $l$  according to [13]:

$$\mathbf{q}_n^{(l)} := (\delta^-)^{n-1} \mathbf{x}^{(l)} \quad \text{for all } n = 1, 2, \dots, k. \quad (22)$$

By the recursive relation  $\mathbf{q}_{n+1}^{(l+1)} = \mathbf{q}_n^{(l+1)} - \mathbf{q}_n^{(l)}$  for the case  $n = k$  we get  $\mathbf{q}_k^{(l+1)} - \mathbf{q}_k^{(l)} = \sigma^{(l)}(\mathbf{q}_1^{(l)}) \Delta l^k$  from Eq. 21 [13]. We assume that  $\mathbf{q}_{n+1}^{(l+1)} = \sum_{l'=n+1}^{(k)} [\mathbf{q}_{l'}^{(l)}] + \sigma(\mathbf{q}_1^{(l)}) \Delta l^k$ , thus the state space in  $(l+1)$  is given by [13]:

$$\mathbf{q}_n^{(l+1)} = \mathbf{q}_n^{(l)} + \mathbf{q}_{n+1}^{(l+1)} = \mathbf{q}_n^{(l)} + \sum_{l'=n+1}^k [\mathbf{q}_{l'}^{(l)}] + \sigma(\mathbf{q}_1^{(l)}) \Delta l^k = \quad (23)$$

$$= \sum_{l'=n}^k [\mathbf{q}_{l'}^{(l)}] + \sigma(\mathbf{q}_1^{(l)}) \Delta l^k. \quad (24)$$

Finally, we obtain the state space dimension for a  $\mathcal{C}^k$ -differentiable neural net. Considering Eq. 23, we can observe that a layer with  $d$ -perceptrons, i.e. a mapping which sends  $\mathbf{x}^{(l)} \mapsto \mathbb{R}^d$ , with a *smoothness of  $k$ -th order* can be represented by the state space form  $\mathbf{q}^{(l)} := [\mathbf{q}_1^{(l)}, \mathbf{q}_2^{(l)}, \dots, \mathbf{q}_k^{(l)}]$ . Thus, it maps into  $\mathbb{R}^{k \cdot d}$ . The more general formula for  $\mathcal{C}^k$ -neural nets is given by [13]:

$$\mathbf{q}_n^{(l+1)} = \sum_{l'=n}^k [\mathbf{q}_{l'}^{(l)}] + \sigma(\mathbf{q}_1^{(l)}) \Delta l^k, \quad \text{for all } n = 1, 2, \dots, k. \quad (25)$$

A neural net with  $k$ -state variables is transformed by a shared activation function  $\sigma^{(l)}$ , which has a  $(d \times d)$ -parameter matrix, but according to the state space theory it is mapped to  $(k \cdot d \times k \cdot d)$ , which means a reduction of the parameter number by the factor  $k^2$  [13]. We use these nets to keep the number of parameters in our neural net as small as possible and to compare the optimal state space dimension instead of the optimal embedding dimension.

### 4.3 Invertible Architecture

The group  $\text{GL}(n, \mathbb{R})$  is the group of invertible matrices, thus we have to equip our neural net with an invertible architecture. The input vector is given by  $\varphi^{(l)} := \mathbf{x}^{(l)} = (\mathbf{u}_1^{(l)}, \mathbf{u}_1^{(l)})$ , splitting the  $l$ -th layer into  $\mathbf{u}_1^{(l)}$  and  $\mathbf{u}_2^{(l)}$  [8,28]:

$$\mathbf{u}_1^{(l)} = \left( \mathbf{v}_1^{(l)} - \psi_2(\mathbf{u}_2^{(l)}) \right) \cdot \exp\left(-\xi_2(\mathbf{u}_2^{(l)})\right), \quad (26)$$

$$\mathbf{u}_2^{(l)} = \left( \mathbf{v}_2^{(l)} - \psi_2(\mathbf{v}_1^{(l)}) \right) \cdot \exp\left(-\xi_1(\mathbf{v}_1^{(l)})\right). \quad (27)$$

The two input vectors  $\mathbf{v}_1^{(l)}$  and  $\mathbf{v}_2^{(l)}$  of the layer ensure that the output is trivially invertible. The multiplication is to be understood elementwise, as in Eq. 18:

$$\mathbf{v}_1 = \mathbf{u}_1^{(l)} \cdot \exp\left(\xi_2(\mathbf{u}_2^{(l)}) + \psi_2(\mathbf{u}_2^{(l)})\right), \quad (28)$$

$$\mathbf{v}_2 = \mathbf{u}_1^{(l)} \cdot \exp\left(\xi_2(\mathbf{u}_2^{(l)}) + \psi_2(\mathbf{u}_2^{(l)})\right). \quad (29)$$



**Fig. 3.** Mean confident persistence landscapes of the investigated datasets cifar10 and cifar100 with a resolution of 1000. The mean persistence landscapes were computed on truncated persistence diagrams, with a confidence of  $\alpha = 0.95$  for a representative of a persistent homology class tracked on the filtration.

We operate on the Lie group  $GL(n, \mathbb{R})$ , therefore all layers are invertible – except the very first one, see Fig. 2 and Sect. 5. The functions  $\psi_i, \xi_j : \mathbf{u}^{(l)}(M) \rightarrow (\xi_j \circ \mathbf{u}^{(l)})(M), \mathbf{u}^{(l)} \mapsto \sigma(W^{(l+1)} \cdot \mathbf{u}^{(l)} + \mathbf{b}^{(l+1)})$  are implemented as neural subnets by themselves. These nets do not have to be invertible in general.

## 5 Numerical Experiments

We train an autoencoder and use as input heavily noisy images and map them to their noiseless original. It is not relevant whether the mean persistence landscapes are computed over the noisy data or over the originals, since the primer can deal with noise. In principle, the persistence landscapes should be calculated for the data manifold. We use invertible  $\mathcal{C}^1$ -nets and isolate the innermost layer of the autoencoder, see Sect. 4.2 and Fig. 2. To test our prediction for the embedding dimension, we parameterize the innermost layer using a multiple of 2 perceptrons. The nets are trained and evaluated according to Fig. 4.

*Persistence Landscapes* The persistence landscapes are used for prediction, computed on the persistence diagrams of the datasets cifar10 and cifar100 [17], see Fig. 3. We use the  $\alpha$ -complex as filtration, see Melodia et al. [20]. The maximum  $\alpha$ -square is set to  $e^{12}$ . The maximal expansion dimension of the simplicial complex is 10. The maximal edge length is set to 0.1. All curves in the mean persistent landscapes are smoothened by a Gaussian filter with  $\sigma = 2$ .


*Neural Nets* The data manifold is supposed to be some commutative Lie group. We perturb each input vector  $\mathbf{x}^{(0)}$  with Gaussian noise, such that the input is

**Table 1.** The representatives from the mean confidence persistent landscape are counted per homology group. We give the minimal embedding dimension for some compact object on a connected commutative Lie group, having approximately as many representatives of the respective homology groups. Further, we give an embedding dimension for an object consisting of all such representatives. We apply Eq. 25 to compute the required state space dimension of resnets with different degrees of smoothness.

Data \ Features	Homology groups					$\approx$ Embedding dim.				
	$H_0$	$H_1$	$H_2$	$H_3$	$H_4$	$U$	$\mathcal{C}^1$	$\mathcal{C}^2$	$\mathcal{C}^3$	$\mathcal{C}^4$
cifar10	12	16	40	59	50	$184 \pm 44$	x	x	x	x
cifar100	13	18	34	46	48		x	x	x	x
Solutions of	12	16	$9 \pm 4$	$8 \pm 3$	$7 \pm 15$					
Eq. 16 & Eq. 17	13	18	$9 \pm 2$	$8 \pm 10$	$7 \pm 13$					

weighted  $0.8 \cdot \mathbf{x}^{(0)} + 0.5 \cdot \epsilon$  with  $\epsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ . These input vectors are trained to the original vector  $\mathbf{x}^{(0)}$ . We use  $L(\mathbf{x}^{(0)}, \mathbf{y}) = \frac{1}{n} \sum_{i=1}^n \left( \mathbf{x}_i^{(0)} - \mathbf{y}_i \right)^2$  as loss function. The sigmoid activation function  $\sigma(\mathbf{x}^{(l)}) = \frac{1}{1+e^{-\mathbf{x}^{(l)}}}$  is applied as elementwise activation function throughout the net. In all experiments dense neural nets are used with a 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.  $\varphi^{(0)}(x^{(0)}) = \sigma \left( \sum_{i=1}^n \mathbf{W}_{ji}^{(1)} \mathbf{x}_i^{(0)} + \mathbf{b}_i^{(1)} \right)$ . It transfers the data to the desired embedding dimension. The subsequent layers are implemented as  $\mathcal{C}^1$ -residual invertible layers, see Sect. 4.2 and Sect. 4.3. A total of 5 of these hidden layers are used, all with the same embedding dimension. More details can be found in Appx. C.

## 5.1 Results

Tab. 1 shows the representatives of homology classes from their respective groups for cifar10 and cifar100. Further, it shows the estimated minimal embedding dimension for a manifold with these homology groups for  $\mathcal{C}^1$  to  $\mathcal{C}^4$ -resnets. The representatives for the  $k$ -th homology group  $\sharp_k^H \left( \bar{\mathfrak{L}}(\mathfrak{P}(K(X))) \right)$  of the respective dataset are counted. Then, we seek for an embedding dimension to represent a manifold with the number of spheres  $q$  with these homology groups. A Monte Carlo approach is used to find integer solutions of Eq. 17. All models below our dimensional threshold show explosions of the gradient, see Fig. 4. The dimension of the respective loss function can be identified by the transition , from the lowest tested dimension to the highest, see Fig. 4. The models above the threshold remain stable. Higher parameterized tested models are stable. A single net still shows an explosive gradient in Fig. 4 (b). This could be due to statistical errors in parametrization by mean confidence persistence landscapes. The net showing the deflection of gradient is close to our threshold. We specify the parameterization



**Fig. 4.** The experiments show the mean squared error loss function on the validation dataset. We trained a total of 392 neural nets ten times per dataset. Each line represents one neural net as an averaged loss function. The architecture can be found in Sect. 4.3, Appx. C and Fig. 2. The dimension of the net is chosen as a multiple of 2, i.e. 2, 4, 6,  $\dots$ , 784. The  $\bullet$ -marked loss functions are parameterized above the determined threshold from Tab. 1. The gray loss functions are below the defined threshold.

for an embedding as well as for neural nets of the smoothness degree 1–4. The neural net is capable of eliminating the additional structure of noise, which is no longer part of the abelian Lie group.

## 6 Conclusions

Based on the theory of Lie groups and persistent homology a method for invertible resnets was developed to parameterize them with good precision due to the assumption that the data lies on or near by some commutative Lie group. This forms an approximate solution for a special case of learning problems, allowing to restrict the domain to a commutative group. The initial problem could be simplified by decomposing the Lie group into a product space  $\mathbb{R}^p \times \mathbb{T}^q$ . The homology groups of the  $q$ -torus and the Euclidean space are well known. The application of fully invertible neural nets to resnets demonstrates our results on a different Lie group from Euclidean space. By applying Künneth’s theorem the homology groups of the topological factor spaces could be connected with the ones of their product space. Using mean confident persistence landscapes the homology groups were estimated. A threshold was computed, see sect. 5, which is determined by the state space dimension of the resnets according to a representation up to homology equivalence. The mean persistence landscapes are particularly useful for short-lived isolated collections of data points, which in particular generate elements from homology classes at the beginning of the filtering and become visible at the lower left corner of the persistence landscapes, see Fig. 3 (a). The theoretical work was tested with numerical experiments on

toy data. We do not limit the processing of the neural net within the layers to an abelian group. Instead, we assume that the net learns this structure from the input domain. This seems to hold, according to our experiments. Still, it can cause inaccuracies in the estimation of the dimensional threshold. The abelianization of the invertible layers would solve this problem.

*Acknowledgements* The work is partially supported by Siemens Power and Gas. The authors would like to thank Noah Becker and David Haller for proofreading.

*Code* The code can be found at: <https://github.com/karhunenloeve/PHIA>.

## References

1. Amari, S.: Information geometry of the EM and em algorithms for neural networks. *Neural Networks* **8**(9), 1379–1408 (1995). [https://doi.org/10.1016/0893-6080\(95\)00003-8](https://doi.org/10.1016/0893-6080(95)00003-8)
2. Amari, S.: Neural learning in structured parameter spaces - natural riemannian gradient. In: *Advances in Neural Information Processing Systems* 9. pp. 127–133 (1996). <https://doi.org/10.5555/2998981.2998999>
3. Ardizzone, L., Kruse, J., Rother, C., Köthe, U.: Analyzing inverse problems with invertible neural networks. In: *7th International Conference on Learning Representations* (2019)
4. Boissonnat, J.D., Chazal, F., Yvinec, M.: *Geometric and Topological Inference*. Cambridge University Press (2018). <https://doi.org/10.1017/9781108297806.014>
5. Bubenik, P.: Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research* **16**, 77–102 (2015). <https://doi.org/10.1007/s10208-014-9201-4>
6. Bubenik, P., Dlotko, P.: A persistence landscapes toolbox for topological statistics. *Journal of Symbolic Computation* **78**, 91–114 (2017). <https://doi.org/10.1016/j.jsc.2016.03.009>
7. Cartan, E.: *La théorie des groupes finis et continus et l’analyse situs*. Gauthier–Villars (1952)
8. Dinh, L., Sohl-Dickstein, J., Bengio, S.: Density estimation using real NVP. In: *5th International Conference on Learning Representations* (2017)
9. Edelsbrunner, H., Harer, J.: Persistent homology – a survey. *Contemporary mathematics* **453**, 257–282 (2008). <https://doi.org/10.1090/conm/453/08802>
10. Fasy, B., Lecci, F., Rinaldo, A., Wasserman, L., Balakrishnan, S., Singh, A., et al.: Confidence sets for persistence diagrams. *The Annals of Statistics* **42**(6) (2014). <https://doi.org/10.1214/14-AOS1252SUPP>
11. Fukumizu, K., Amari, S.: Local minima and plateaus in hierarchical structures of multilayer perceptrons. *Neural Networks* **13**(3), 317–327 (2000). [https://doi.org/10.1016/S0893-6080\(00\)00009-5](https://doi.org/10.1016/S0893-6080(00)00009-5)
12. Hatcher, A.: *Algebraic Topology*. Cambridge University Press (2002). <https://doi.org/10.1017/S0013091503214620>
13. Hauser, M., Gunn, S., Saab, S., Ray, A.: State-space representations of deep neural networks. *Neural Computation* **31**(3) (2019). <https://doi.org/10.1162/neco.a.01165>

14. Hauser, M., Ray, A.: Principles of riemannian geometry in neural networks. In: Advances in Neural Information Processing Systems 30. pp. 2807–2816 (2017). <https://doi.org/10.5555/3294996.3295040>
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016). <https://doi.org/10.1109/CVPR.2016.90>
16. Johnson, J.: Deep, skinny neural networks are not universal approximators. In: 7th International Conference on Learning Representations (2019)
17. Krizhevsky, A.: Learning multiple layers of features from tiny images (2009)
18. Lee, J.: Introduction to smooth manifolds. Springer (2013). <https://doi.org/10.1007/978-1-4419-9982-5>
19. Lin, H., Jegelka, S.: Resnet with one-neuron hidden layers is a universal approximator. In: Advances in Neural Information Processing Systems 31. pp. 6172–6181 (2018). <https://doi.org/10.5555/3327345.3327515>
20. Melodia, L., Lenz, R.: Persistent homology as stopping-criterion for voronoi interpolation. To appear in the International Workshop on Combinatorial Image Analysis (2020)
21. Nguyen, T., Ardizzzone, L., Köthe, U.: Training invertible neural networks as autoencoders. In: Pattern Recognition. pp. 442–455 (2019). [https://doi.org/10.1007/978-3-030-33676-9\\_31](https://doi.org/10.1007/978-3-030-33676-9_31)
22. Rubinstein, B.I.P., Rubinstein, J.H.: A geometric approach to sample compression. Journal of Machine Learning Research **13**, 1221–1261 (2012). <https://doi.org/10.5555/2188385.2343686>
23. Rumelhart, D., Hinton, G., Williams, R.: Learning representations by back-propagating errors. Nature (1986). <https://doi.org/10.1049/esn.1986.0002>
24. Stephan, K., Matthias, K.: A quick proof of the rational hurewicz theorem and a computation of the rational homotopy groups of spheres. Mathematical Proceedings of the Cambridge Philosophical Society **136**(3), 617–623 (2004). <https://doi.org/10.1017/S0305004103007114>
25. Yang, H., Amari, S.: The efficiency and the robustness of natural gradient descent learning rule. In: Advances in Neural Information Processing Systems 10. pp. 385–391 (1997). <https://doi.org/10.5555/3008904.3008959>
26. Yang, L., Yu, W.: Backpropagation with homotopy. Neural Computation **5**(3), 363–366 (1993). <https://doi.org/10.1162/neco.1993.5.3.363>



## A Supplementary Material for Sect. 2

### A.1 Groups

**Definition A.1.** A set  $G$  together with a map  $\mu : G \times G \rightarrow G$ ,  $(x, y) \mapsto xy$  and an element  $e = e_G$  is called group if the following conditions are fulfilled:

- (1)  $(xy)z = x(yz)$  for all  $x, y, z \in G$ , thus the elements are associative,
- (2)  $xe = ex = x$  for all  $x \in G$ , so there is a neutral element and
- (3)  $\forall x \in G \exists x^{-1} \in G$ , such that  $xx^{-1} = x^{-1}x = e$ .

**Definition A.2.** [35, p. 196] Let  $G$  be a group and contain a subset  $H \subseteq G$ .  $G$  is called free group with basis  $B$ , if for each  $H$  and each function  $\psi : B \rightarrow H$  a unique homomorphism  $\varphi : G \rightarrow H$  with  $\varphi(x) = \psi(x)$  exists for all  $x \in B$ .

*Example A.2.1.* The most prominent example for a free group are the integers with addition  $(\mathbb{Z}, +)$  and basis  $B = \{1\}$ . Consider  $\psi : B \rightarrow G$  with  $g = \psi(1)$  and  $g \in G$  and  $\varphi(x) = g^n$  for each  $n \in \mathbb{Z}$ . Then  $\varphi$  is the only group homomorphism of  $(\mathbb{Z}, +)$  into  $G$ , such that  $\varphi(1) = \psi(1)$ .

**Definition A.3.** Let  $\mathbf{X} = \{X, \varphi_i\}_{i \in I}$  and  $\mathbf{Y} = \{Y, \psi_i\}_{i \in I}$  be two algebraic structures of the same type  $\sigma = \{n_i\}_{i \in I}$ , such that for each  $i \in I$  and  $n_i \in \mathbb{N}$ , the arity of the fundamental operations is denoted with  $\varphi_i$  and  $\psi_i$ . The projection  $\varphi : \mathbf{X} \rightarrow \mathbf{Y}$  is called homomorphism, if for all  $x_i \in X$

$$\varphi(f_i(x_1, \dots, x_n)) = g_i(\varphi(x_1), \dots, \varphi(x_n)). \quad (30)$$

*Example A.3.1.* The function  $\varphi : \mathbb{Z} \rightarrow \mathbb{Z}/2\mathbb{Z}$  is a homomorphism of groups. The group  $(\mathbb{Z}/2\mathbb{Z}, +)$  is the quotient space that assembles all even integers to zero and all odd integers to one, where the respective equivalence class is defined.

*Remark A.3.1.1.* Recalling group homomomorphisms and isomorphisms helps to gain proper intuition on chain and boundary groups. Isomorphisms from one algebraic structure to another exist if these two structures are considered to be essentially ‘equal’. A homomorphism is weaker. It preserves ‘some’ of the structure. Since isomorphisms are homomorphisms, this preserved structure can be regarded as a kind of ‘subset’. Homomorphisms preserve operations, but allow elements of one algebraic structure to collapse into some elements of the other.

**Definition A.4.** A subgroup  $H \subseteq G$  is a group if a homomorphism  $\varphi : G \rightarrow H$  exists, which maps  $\varphi(e_G)$  to  $e_H$  and  $\varphi(xy)$  to  $\varphi(x)\varphi(y)$  for all  $x, y \in G$ . A subgroup is called *normal* if  $xHx^{-1} = H$ , for all  $x \in G$ .

*Example A.4.1.* [29, p. 5] Consider  $\varphi : G \rightarrow H$  as a surjective group homomorphism. Then  $K := \ker \varphi$  is a normal subgroup of  $G$ . There is a unique map  $\bar{\varphi} : G/K \rightarrow H$ , which yields  $\bar{\varphi} \circ \pi = \varphi$ , with the canonical projection  $\pi$ . The *factor map*  $\bar{\varphi}$  is an isomorphism of groups.

*Proof.* Let  $x \in G$  and  $k \in K$ . Then  $\varphi(xkx^{-1}) = \varphi(x)\varphi(k)\varphi(x^{-1}) = \varphi(x)e_H\varphi(x)^{-1} = e_H$ , thus  $xkx^{-1}$  is an element of  $\ker \varphi = K$ . It follows that  $xKx^{-1} \subset K$ . Due to the properties of the neutral element  $x^{-1}Kx \subset K$ , i.e.  $K \subset xKx^{-1}$ , we get  $K = xKx^{-1}$ .  $\square$

*Remark A.4.1.1.* Similar as subsets behave to sets, *subgroups* behave to groups.

## A.2 Lie groups

**Definition A.5.** Let  $G$  be a commutative abelian Lie group. The exponential map  $\exp := \exp_G : T_e G \rightarrow G$  maps from the tangent space  $T_e G$  in  $e$  into the group  $G$ . It also maps in every point into  $G$  due to homogeneity. It is defined as  $\exp(\mathfrak{X}) = \alpha_{\mathfrak{X}}(1)$  for a left invariant vector field  $\mathfrak{X} \in T_e G$  and a maximal integral curve  $\alpha_{\mathfrak{X}}$  with a starting point  $e$ . For all  $s, t \in \mathbb{R}$  and each  $\mathfrak{X} \in T_e G$  the exponential map fulfils the properties

- (a)  $\exp(s\mathfrak{X}) = \alpha_{\mathfrak{X}}(s)$ ,
- (b)  $\exp(s+t)\mathfrak{X} = \exp(s\mathfrak{X})\exp(t\mathfrak{X})$ .

*Proof.* Let  $\gamma(t) = \alpha_{\mathfrak{X}}(st)$  be a curve. Then  $\gamma(0) = e$  and

$$\frac{d}{dt}\gamma(t) = s\dot{\alpha}_{\mathfrak{X}}(st) = sv_{\mathfrak{X}}(\alpha_{\mathfrak{X}}(st)) = v_{s\mathfrak{X}}(\gamma(t)). \quad (31)$$

Let  $\mathfrak{X} \in T_e G$ , then the integral curve  $\alpha_{\mathfrak{X}}$  is defined on  $\mathbb{R}$ . Let  $y \in G$  and  $\alpha_1(t) = y\alpha(t)$ . Differentiating with respect to  $t$  yields

$$\frac{d}{dt}\alpha_1(t) = T_{\alpha(t)}y \frac{d}{dt}\alpha(t) = T_{\alpha(t)}yv_{\mathfrak{X}}(\alpha(t)) = v_{\mathfrak{X}}(\alpha_1(t)) \quad (32)$$

by the left invariance of  $v_{\mathfrak{X}}$ . So  $\alpha_1$  is also an integral curve for  $v_{\mathfrak{X}}$ . Let  $I$  be the domain of  $\alpha_{\mathfrak{X}}$  and  $t_1 \in I$  with  $x_1 = \alpha_{\mathfrak{X}}(t_1)$ . Then  $\alpha_1(t) := x_1\alpha_{\mathfrak{X}}(t)$  is an integral curve for  $v_{\mathfrak{X}}$  with starting point  $x_1$  given by  $\alpha_2 : t \mapsto \alpha_{\mathfrak{X}}(t + t_1)$ . The integral curve has domain  $I - t_1$ . We conclude that  $I \subset I - t_1$ . It follows that  $s + t_1 \in I$  for all  $s, t_1 \in I$  so  $I$  must be equal to  $\mathbb{R}$ .

For a fixed  $x \in \mathbb{R}$  we get  $\gamma : t \mapsto \alpha_{\mathfrak{X}}(s)\alpha_{\mathfrak{X}}(t)$  as maximal integral curve for  $v_{\mathfrak{X}}$  with starting point  $\alpha_{\mathfrak{X}}(s)$ . On the other hand, the same applies to  $\epsilon : t \mapsto \alpha_{\mathfrak{X}}(s+t)$ . Thus  $\gamma = \epsilon$  and we get

$$\exp((s+t)\mathfrak{X}) = \alpha_{\mathfrak{X}}(s+t) = \alpha_{\mathfrak{X}}(s)\alpha_{\mathfrak{X}}(t) = \exp(s\mathfrak{X})\exp(t\mathfrak{X}). \quad (33)$$

□

*Remark A.5.1.* The exponential map is a smooth map  $\exp_G : T_e G \rightarrow G$  and a local diffeomorphism around 0.

A Lie group  $G$  is also a smooth manifold, as defined in Sect. 2.2. Its tangential space is a linear space, which we denote as  $\mathfrak{g}$ .

**Definition A.6.** Each linear space  $\mathfrak{g}$  equipped with a bilinear form  $[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$  is called Lie algebra, if the bilinear form, also called Lie bracket, is

- (1) antisymmetric, i.e.  $[\mathfrak{X}, \mathfrak{Y}] = -[\mathfrak{Y}, \mathfrak{X}]$  and the
- (2) Jacobi identity holds  $[\mathfrak{X}, [\mathfrak{Y}, \mathfrak{Z}]] = [\mathfrak{Y}, [\mathfrak{Z}, \mathfrak{X}]] + [\mathfrak{Z}, [\mathfrak{X}, \mathfrak{Y}]]$ , for all  $\mathfrak{X}, \mathfrak{Y}, \mathfrak{Z} \in \mathfrak{g}$ .

*Remark A.6.1.* For a Lie group  $G$  with subgroup  $H$  both properties are equivalent [29, pp. 25-27].

**Proposition A.1.** [29, p.17] Let  $G$  be a Lie group and  $H$  a subgroup. Then the following properties are equivalent:

- (1) There is an open neighborhood  $U$  of  $e$  such that  $U \cap H = \{e\}$ .
- (2) The group  $H$  is discrete.
- (3) For each compact subset  $C \subset G$  the intersection  $C \cap H$  is finite.
- (4) The group  $H$  is a closed Lie subgroup with Lie algebra  $\{0\}$ .

*Proof.* Take any element  $h \in H$ . Then  $hU$  is an open neighbourhood of  $h$  in  $G$ .  $hU \cap H = h(U \cap h^{-1}H)$  and from the group structure of  $H$  it follows that  $h(U \cap h^{-1}H) = h(U \cap H) = \{h\}$ .

Choose  $U$  as an open neighborhood centered around  $e = e_G$ , such that  $U \cap H = \{e\}$ . Let  $g \in G$  be a point in the closure  $\overline{H}$ . Then there is a sequence  $\{h_i\}$  in  $H$  which converges to  $g$ . It follows that  $h_{i+1}h_i^{-1} \rightarrow gg^{-1} = e$ , with  $i \rightarrow \infty$ . For an  $i$  large enough,  $h_{i+1}h_i^{-1} \in U \cap H = \{e\}$ , i.e.  $h_i = h_{i+1}$ . The sequence  $h_i$  becomes stationary after a certain index. Therefore  $h_i = g$  for some  $i$ . We conclude that  $g \in H$  for any  $g$ . Further we see that  $\overline{H} \subset H$  and therefore  $H$  is closed. Since  $H$  is closed, the intersection  $H \cap C$  is also closed, thus compact. For a  $c \in H \cap C$  we choose an open neighborhood  $cU$  in  $G$ , such that  $cU \cap H = \{c\}$ . Then  $\{cU \mid c \in H \cap C\}$  is an open cover of  $H \cap C$  with no proper subcover. Since  $H \cap C$  is compact, the cover must be finite and thus  $H \cap C$  is finite.

Let  $g \in G$  a point in the closure  $\overline{H}$ . We choose a compact neighborhood  $C$  around  $g$ , such that  $g$  lies in the closure  $\overline{H \cap C}$ . The latter is finite, thus closed. Therefore  $g$  is in the closure  $\overline{H \cap C} \subset H$  and contained in  $H$ . We conclude that  $H$  is a closed Lie subgroup. Let  $\mathfrak{g}$  be the Lie algebra of  $H$  consisting of  $\mathfrak{X} \in \mathfrak{g}$  with  $\exp(\mathbb{R}\mathfrak{X}) \subset H$ . A local diffeomorphism  $\exp : \mathfrak{g} \rightarrow G$  to 0 gives an open neighborhood  $\mathfrak{o}$  around 0 in  $\mathfrak{g}$ . The exponential function is injective on  $\mathfrak{o}$ . If we take  $\mathfrak{X} \in \mathfrak{g} \setminus \{0\}$ , then there is an  $\epsilon > 0$ , such that  $[-\epsilon, \epsilon]\mathfrak{X} \subset \mathfrak{o}$ . We define a curve  $\gamma : [-\epsilon, \epsilon] \rightarrow G$ ,  $t \mapsto \exp(t\mathfrak{X})$ , with compact image. Thus  $\text{im}(\gamma) \cap H$  and  $\{t \in [-\epsilon, \epsilon] \mid \exp(t\mathfrak{X}) \in H\}$  are finite. Hence,  $\mathfrak{X} \notin \mathfrak{h}$  and  $\mathfrak{h} = \{0\}$ .

$H$  is a closed smooth submanifold of  $\mathfrak{X}$ , where  $\dim H = 0$ . It follows that an open neighborhood  $U$  exists from  $e$  to  $G$ , such that  $U \cap H = \{e\}$ .  $\square$

### A.3 Decomposition of Connected Commutative Lie groups

**Theorem A.1.** [39, 29, p. 116, p. 26] Every connected commutative (or in other words abelian) Lie group  $G$  is isomorphic to some product space  $\mathbb{R}^p \times \mathbb{T}^q$  with  $\dim G = p + q$ , i.e.  $G \simeq \mathbb{R}^p \times \mathbb{T}^q$ .

*Proof.* Choosing  $G$  as a connected abelian Lie group, the Lie algebra  $\mathfrak{g}$  is commutative, such that  $[\mathfrak{X}, \mathfrak{Y}] = 0$  for all  $\mathfrak{X}, \mathfrak{Y} \in \mathfrak{g}$ . It follows that the exponential map  $\exp(\mathfrak{X})\exp(\mathfrak{Y}) = \exp(\mathfrak{X} + \mathfrak{Y})$  is given for all  $\mathfrak{X}, \mathfrak{Y} \in \mathfrak{g}$ . Thus  $\exp : \mathfrak{g} \rightarrow G$  is a Lie group homomorphism from  $(\mathfrak{g}, 0, +)$  to  $G$ . It follows that  $\exp(\mathfrak{g})$  is a subgroup of  $G$ , thus  $\exp(\mathfrak{g}) = G_e$ . Since  $G$  is connected,  $G = G_e$ . Therefore  $\text{im } \exp(\mathfrak{g}) = G$  and  $\exp : \mathfrak{g} \rightarrow G$  is a surjective Lie group homomorphism. If  $\Omega$  is the closed subgroup  $\ker(\exp)$  of  $\mathfrak{g}$ , then according to the Isomorphism Theorem for groups:  $G \simeq \mathfrak{g}/\Omega$  [29, pp. 5-7]. Since  $\exp$  is a local diffeomorphism at 0, there

is an open neighborhood  $\Gamma$  of 0 in  $\mathfrak{g}$  for which  $\exp$  is an injective map. It follows that  $\Gamma \cap \Omega = \{0\}$ . Thus  $\Omega$  is a discrete subgroup of  $\mathfrak{g}$ .

We need a corresponding basis for a discrete group consisting of finite-dimensional linear independent elements with coefficients in  $\mathbb{Z}$ .

**Lemma A.1.1.** [29, p. 27] A subset of a finite dimensional linear space  $V$  can be written by linear independent components  $\omega_1, \dots, \omega_p$  of  $V$  as a direct sum with coefficients in  $\mathbb{Z}$ , such that  $\Omega = \mathbb{Z}\omega_1 \oplus \dots \oplus \mathbb{Z}\omega_p$ .

*Proof.* We show that by induction and start with  $\dim V = 1$ . Choosing a basis, we identify  $V$  with  $\mathbb{R}$ . Then  $\Omega$  is a discrete subgroup of  $\mathbb{R}$ . We assume that  $\Omega \neq \{0\}$ , then there is an element  $x \in \Omega \setminus \{0\}$  with  $x > 0$ . The section  $\Omega \cap [0, x]$  is finite and therefore contains a smallest element  $\omega$ . The section  $\Omega \cap ]0, 1[\omega$  is equal to the empty set, thus  $\mathbb{Z}\omega \subset \Omega$ . Next we consider an element  $y \in \Omega$ , then  $y \notin \mathbb{Z}\omega$  would mean that  $y \in ]n, n+1[\omega$  for a corresponding  $n \in \mathbb{Z}$ . This means that  $y - n\omega \in \Omega \cap ]0, 1[\omega = \emptyset$ . This in turn implies  $\Omega \subset \mathbb{Z}\omega$ . Thus  $\Omega = \mathbb{Z}\omega$ .

For the general case we use the idea from van den Ban's proof [29, p.27]. We assume that  $\dim V = n$  for any  $n > 1$ . In the trivial case  $p = 0$  and  $\Omega = \{0\}$  holds. For  $\omega \in \Omega \setminus \{0\}$  the intersection  $\mathbb{R}\omega \cap \Omega$  is discrete in  $\mathbb{R}\omega$ , i.e. an integer multiple  $\mathbb{Z}\omega_1$ . We choose a linear subspace  $U \subset V$ , in order that  $\mathbb{R}\omega_1 \oplus U = V$ , with the canonical projection  $\pi : U \rightarrow V$  and  $C \subset U$  as compact subset:

$$\pi^{-1}(C) = \mathbb{R}\omega_1 + C = ([0, 1]\omega_1 + C) + \mathbb{Z}\omega_1, \quad (34)$$

$$\implies \pi(\Omega) \cap C = \pi((C + [0, 1]\omega_1) \cap \Omega + \mathbb{Z}\omega_1) \quad (35)$$

$$= \pi((C + [0, 1]\omega_1) \cap \Omega). \quad (36)$$

The set  $C$  is compact, so is  $C + [0, 1]\omega_1$ . It follows that the intersection  $(C + [0, 1]\omega_1) \cap \Omega$  is also compact. Thus  $C$  is a discrete compact subset of  $U$ . By induction there are linearly independent elements  $\bar{\omega}_2, \dots, \bar{\omega}_p$  of  $\pi(\Omega)$ , such that  $\pi(\Omega) = \mathbb{Z}\bar{\omega}_2 \oplus \dots \oplus \mathbb{Z}\bar{\omega}_p$ . We choose  $\pi(\omega_i) = \bar{\omega}_i$ , then the elements  $\omega_1, \dots, \omega_p$  are linearly independent and we get the basis  $\Omega = \mathbb{Z}\omega_1 \oplus \dots \oplus \mathbb{Z}\omega_p$ .  $\square$

Since  $\mathfrak{g}$  is a linear space, we can extend the basis  $\mathbb{Z}\omega_1 \oplus \dots \oplus \mathbb{Z}\omega_n$  to the whole Lie algebra  $\mathfrak{g}$ . Thus we get a linear isomorphism  $\varphi : \mathfrak{g} \rightarrow \mathbb{R}^n$ , where  $\mathbb{R}^n \simeq \mathbb{R}^p \times \mathbb{R}^q$  with  $p + q = n$ . The map  $\exp \circ \varphi^{-1} : \mathbb{R}^n \rightarrow G$  is a surjective Lie group homomorphism, i.e. a group homomorphism that is smooth, with smooth inverse. According to the Isomorphism Theorem for groups, this group also factorizes by a bijective group homomorphism [29, pp. 5-7]:

$$\begin{array}{ccccc}
 & & \exp & & \\
 & & \curvearrowright & & \\
 \mathfrak{g} & \xrightarrow{\varphi} & \mathbb{R}^p \times \mathbb{R}^q & \xrightarrow{\exp \circ \varphi^{-1}} & G \\
 & \searrow \pi \circ \varphi & \downarrow \pi & \nearrow \exp \circ \varphi^{-1} & \\
 & & (\mathbb{R})^p \times \mathbb{R}/\mathbb{Z}^q & & 
 \end{array}$$

The canonical map  $\pi : \mathbb{R}^n \simeq \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}^p \times (\mathbb{R}/\mathbb{Z})^q$  is locally diffeomorphic and  $\exp \circ \varphi^{-1} \circ \pi$  as well. Due to the homogeneity of commutative lie groups,

the map is everywhere a diffeomorphism. The projection  $\overline{\exp \circ \varphi^{-1}}$  is bijective, thus also a diffeomorphism and thus an isomorphism of Lie groups. By definition  $\mathbb{R}^q/\mathbb{Z}^q \simeq (\mathbb{R}/\mathbb{Z})^q \simeq \mathbb{T}^q$  and the proof follows.  $\square$

## B Supplementary Material for Sect. 3

### B.1 Properties of Simplicial Complexes

**Definition B.1.** A simplicial complex  $K$  is *connected* if it cannot be composed of a disjoint union of two or more non-empty subcomplexes.

**Definition B.2.** An embedded simplicial complex is *path-connected* if there is a path from any vertex to any other as a composition of 1-simplices.

**Proposition B.1.** A simplicial complex  $K$  is path-connected if it is connected and vice versa. Thus the notions are equivalent for simplicial complexes.

*Proof.* Assume  $K$  is path-connected and not connected. Choose two subcomplexes  $U, V \subset K$  with  $U \cup V = K$ . Choose a vertex  $u_0 \in U$  and  $v_0 \in V$ . Thus a path from  $u_0$  to  $v_0$  exists, due to the path-connectedness of the simplicial complex. If  $u_i$  is the last vertex in  $U$ , then the vertex  $u_{i+1}$  can be contained neither in  $U$  nor in  $V$ , due to the disjoint decomposition of the simplicial complex. This contradicts the assumption that it is path-connected.

Consider two vertices  $u, v \in K$  that are not path-connected. Define  $U, V \subset K$  as path-connected subcomplexes containing  $u \in U$  and  $v \in V$ . Choose  $w \in U \cap V \neq \emptyset$ , then there is a path from  $u$  to  $w$  and also from  $w$  to  $v$ . A composition of paths gives a path from  $u$  to  $v$ , which contradicts the assumption that  $u$  and  $v$  are not path-connected. Thus  $U \cap V = \emptyset$  and  $K$  is not connected.  $\square$

### B.2 Direct Sums of Homology Groups

**Theorem B.1.** [34, p. 109] The  $k$ -th simplicial homology group  $H_n^\Delta(K(X))$  of a simplicial complex  $K$  for a field  $\mathbb{F}$  is isomorphic to the direct sum of the homology groups of all subcomplexes on its filtration, i.e.  $H_n^\Delta(K(X)) \simeq \bigoplus_{i \in I} H_n^\Delta(K^{\omega_i}(X))$ .

*Proof.* Let  $\{K^{\omega_i}\}_{i \in I}$  be the set of all connected components of the simplicial complex  $K(X)$  of a point set  $X$  and  $H_k^\Delta(K(X))$  its simplicial homology groups. We denote the homology groups of the components with  $H_k^\Delta(K^{\omega_i}(X))$ . With  $C_k(K^{\omega_i}(X))$ , we denote the group of  $k$ -chains for the connected components of the simplicial complex. Then  $C_k(K^{\omega_i}(X)) \subseteq C_k(K(X))$  is a subgroup for all  $i$ .

**Lemma B.1.1.** The group of  $k$ -chains can be written as a direct sum of the connected components, i.e.  $C_k(K(X)) \simeq \bigoplus_{i \in I} C_k(K^{\omega_i}(X))$ . This also applies to the lower chain groups.

*Proof.* Using the boundary operator, which is a homomorphism of groups, we get  $C_{k-1}(K(X)) \simeq \bigoplus_{i \in I} \partial_k(C_k(K^{\omega_i}(X)))$ . An element in the  $(k+1)$ -th chain group  $v \in C_{k+1}(K(X))$  is represented by  $v = \sum_{i \in I} v_{\omega_i}$ . Thus  $\partial_{k+1}(v) \in C_k(K(X))$  can be broken down to  $\partial_{k+1}(v) = \sum_{i \in I} \partial_{k+1}(v_{\omega_i})$  where  $v_{\omega_i} \in C_{k+1}^{\omega}(K^{\omega_i}(X))$ .  $\square$

The groups of boundaries can also be decomposed. The kernel of  $C_k(K^{\omega_i}(X))$  can be written as  $\ker \partial_k(C_k(K^{\omega_i}(X))) = \ker \partial_k(C_k(K(X))) \cap C_k(K^{\omega_i}(X))$ .

$$\ker \partial_k \simeq \bigoplus_{i \in I} \partial_k(C_k(K^{\omega_i}(X))), \text{ which implies} \quad (37)$$

$$\Rightarrow \frac{\ker \partial_k(C_k(K(X)))}{\text{im } \partial_{k+1}(C_{k+1}(K(X)))} \simeq \bigoplus_{i \in I} \frac{\ker \partial_k(C_k(K^{\omega_i}(X)))}{\text{im } \partial_{k+1}(C_{k+1}(K^{\omega_i}(X)))} \quad (38)$$

$$\Rightarrow H_k^{\Delta}(K(X)) \simeq \bigoplus_{i \in I} H_k^{\Delta}(K^{\omega_i}(X)). \quad (39)$$

$\square$

*Remark B.1.1.1.* We obtain an isomorphism for the singular homology groups from the homology groups of the set of path components. This can be thought of as a direct sum of all ‘loops’ from every possible path component in space.

**Definition B.3.** A *path component* of a topological space  $X$  is an equivalence class under the equivalence relation  $x \sim y$  if there is a path from  $x$  to  $y$ .

*Remark B.3.1.* Since for simplicial complexes the property to be connected corresponds to path-connectedness, the coordinatewise sum of the connected components is a generator of the whole group.

**Theorem B.2.** [34, p. 129] Let  $X$  be a topological space and  $\{X^{\omega_i}\}_{i \in I}$  a set of path components of  $X$  and  $\iota_{\omega_i} : X^{\omega_i} \hookrightarrow X$  the inclusion. Then the maps  $(\iota_{\omega_i})_{\star} : H_k(X^{\omega_i}) \hookrightarrow H_k(X)$  for each  $k \geq 0$  induce an isomorphism between the direct sum of the homology groups of all simplicial subcomplexes and the homology groups of  $X$ :

$$\bigoplus_{i \in I} H_k(X^{\omega_i}) \rightarrow H_k(X). \quad (40)$$

*Proof.* Consider  $C_k(X^{\omega_i})$  to be the free abelian group with basis  $\{\sigma : \Delta^k \rightarrow X^{\omega_i}\}$  of all singular simplices in  $X^{\omega_i}$ . A generator  $\sigma \in C_k(X^{\omega_i})$  is a simplex in  $X$  composed with the inclusion  $\iota_{\omega_i}$ , i.e.  $(\iota_{\omega_i})_{\star} : C_k(X^{\omega_i}) \rightarrow C_k(X)$ ,  $\sigma \mapsto \iota_{\omega_i} \circ \sigma$ . If we take all  $i \in I$ , we can exhaust  $C_k(X)$ , which results in the isomorphism  $\psi : \bigoplus_{i \in I} H_k(X^{\omega_i}) \rightarrow H_k(X)$  at chain level. Taking an element  $x \in \bigoplus_{i \in I} C_k(X^{\omega_i})$  we get a finite sum of the functions  $\sigma_1, \dots, \sigma_n$  with their coefficients  $z_1, \dots, z_n \in \mathbb{Z}$ , i.e.  $x = \sum_{i=1}^n z_i \sigma_i$  and apply the group homomorphism  $\psi$ , resulting in  $\psi(x) = \sum_{i=1}^n z_i \iota_{\omega_i}(\sigma_i)$ . Given a generator  $\sigma : \Delta^k \rightarrow X$ , we know that  $\Delta^k$  is connected and  $\sigma$  is a continuous function. Thus  $\sigma(\Delta^k)$  is contained in a connected component of  $X$ , namely  $\sigma \in C_k(X^{\omega_j})$  and  $\sigma = \psi(\sigma)$ , thus  $\psi$  is surjective.  $\square$

### B.3 Equivalence of Simplicial and Singular Homology

**Theorem B.3.** [34, p. 128] Let  $k \in \mathbb{Z}$  and  $H_k^\Delta(K(X))$  the simplicial  $k$ -th homology group over a triangulable topological space  $X$  and  $K$  the triangulation. Let  $H_k(X)$  be the respective singular homology group. Then the map  $H_k^\Delta(K(X)) \rightarrow H_k(X)$  is an isomorphism of groups.

*Proof.* We follow the proofs of [34,38, p. 128, pp. 190-192] and start from the perspective of the simplicial homology groups by choosing  $K(X)$  as a simplicial complex and  $K(X^n)$  as the  $n$ -skeleton of  $X$ , which consists of all simplices of dimension  $n$  or less. Thus, with  $K(X^{n-1}) \subset K(X)$  we obtain the following commutative diagram of exact sequences, for which  $\ker(\psi_i) = \text{im}(\psi_{i+1})$ :

$$\begin{array}{ccccccc}
 H_{k+1}^\Delta((K(X^n), K(X^{n-1}))) & \xrightarrow{\psi_i} & H_k^\Delta(K(X^{n-1})) & & & & \\
 \downarrow \sigma_i & & \downarrow \sigma_j & & & & \\
 H_{k+1}((X^n, X^{n-1})) & \xrightarrow{\psi'_i} & H_k(X^{n-1}) & & & & \\
 & & & & \psi_j & & \\
 & & & & \downarrow & & \\
 H_{n-1}^\Delta(K(X^{n-1})) & \xleftarrow{\psi_k} & H_k^\Delta((K(X^n), K(X^{n-1}))) & \xleftarrow{\psi_l} & H_k^\Delta(K(X^n)) & & \\
 \downarrow \sigma_m & & \downarrow \sigma_l & & \downarrow \sigma_k & & \\
 H_{k-1}(X^{n-1}) & \xleftarrow{\psi'_k} & H_k((X^n, X^{n-1})) & \xleftarrow{\psi'_l} & H_k(X^n) & & \\
 & & & & \psi'_j & & 
 \end{array}$$

The quotient space  $X^n/X^{n-1}$  contains only simplices of dimension  $n$ . For  $k \neq n$  the group  $C_k((X^n, X^{n+1}))$  is zero. For  $k = n$ , the group  $C_k((X^n, X^{n+1}))$  is free and commutative with a basis consisting of  $k$ -simplices. The cycles  $\ker \partial_k$  are a subgroup of  $C_k((X^n, X^{n+1}))$  and the boundary group  $\text{im } \partial_{k+1}$  is empty. We conclude that  $H_k^\Delta((K(X^n), K(X^{n-1})))$  has a description like  $C_k((X^n, X^{n-1}))$ .

The continuous maps  $\sigma : \Delta^n \rightarrow X$  into the topological space  $X$  give the map  $\psi : \coprod_i (\Delta_i^n, \Delta_i^{n-1}) \rightarrow (X^n, X^{n-1})$  and for the simplicial quotient spaces  $\psi_* : \coprod_i \Delta_i^n / \coprod_i \Delta_i^{n-1} \rightarrow X^n/X^{n-1}$ . In view of these homeomorphisms, the following isomorphism of homology groups results:

$$H_k(\coprod_i \Delta_i^n / \coprod_i \Delta_i^{n-1}) \simeq H_k(X^n/X^{n-1}). \quad (41)$$

A consequence of the Excision Theorem is the existence of an isomorphism  $H_k((X, V)) \rightarrow H_k(X/V)$  for all good pairs  $(X, V)$  [38, pp. 175-182].

**Theorem B.4.** Given the topological spaces  $U \subset V \subset X$  for which the closure  $\overline{U} \subset V^\circ$  is in the interior of  $V$ , the embedding  $(X - U, V - U) \hookrightarrow (X, V)$  induces an isomorphism  $H_k((X - U, V - U)) \rightarrow H_k((X, V))$  for all  $k \in \mathbb{Z}$ .

*Proof.* For a proof we refer to [38, pp. 175-182].  $\square$

*Remark B.4.1.* Recall that the notation  $(X^n, X^{n-1})$  denotes a topological pair, i.e. a shorthand to indicate an inclusion  $\iota : X^{n-1} \hookrightarrow X^n$ .

*Remark B.4.2.* As a reminder, a good pair  $(X, V)$  is given if there is an open neighborhood  $W \subset X$  with  $V \subset W$  such that this deformation retracts onto  $V$ .

Applying transitivity of group operations yields

$$H_k \left( \prod_i (\Delta_i^n, \Delta_i^{n-1}) \right) \simeq H_k \left( \prod_i \Delta_i^n / \prod_i \Delta_i^{n-1} \right), \quad (42)$$

$$H_k \left( \prod_i \Delta_i^n / \prod_i \Delta_i^{n-1} \right) \simeq H_k((X^n, X^{n-1})), \quad (43)$$

$$H_k((X^n, X^{n-1})) \simeq H_k(X^n / X^{n-1}). \quad (44)$$

We see, that  $H_k((X^n, X^{n-1}))$  is zero, for  $k \neq n$  and a free abelian group, on the basis of the cycles given by the continuous map  $\sigma : \Delta^n \rightarrow X$ . It follows that the map  $H_k^\Delta((K(X^n), K(X^{n-1}))) \rightarrow H_k((X^n, X^{n-1}))$  is an isomorphism.

In the commutative diagram of simplicial and singular homology groups above, the maps  $\sigma_i$  and  $\sigma_m$  are isomorphisms. By induction  $\sigma_j$  and  $\sigma_l$  are also isomorphisms. It remains to be shown that  $\sigma_k$  is also an isomorphism. For this purpose we use the Five Lemma [41], from which follows that  $\sigma_k$  is an isomorphism and therefore the map  $H_k^\Delta(K(X^n)) \rightarrow H_k(X^n)$  is an isomorphism.  $\square$

#### B.4 Singular Homology and Smooth Manifolds

To complete the argument we use ideas from a proof of de Rham's Theorem. De Rham's Theorem manifests the fundamental connection of the analytic and topological properties of a smooth manifold, where the pairing of differential forms and singular chains gives an isomorphism from  $k$ -th de Rham cohomology to  $k$ -th singular homology and an isomorphism into simplicial homology.

**Theorem B.5.** The de Rham cohomology of a smooth, oriented and triangulable manifold  $H_k^{\text{dR}*}(M)$  has an isomorphism to the real homology group  $H_k(K(M); \mathbb{R})$ , e.g. calculated by a singular Čech complex  $K$  on some smooth manifold  $M$  with real coefficients.

*Proof.* For a complete proof we refer to [37, 31, Chap. 18, Sect. V.9].

**Theorem B.6.** The  $k$ -th smooth singular homology group is defined as

$$H_k^\infty(M) = \frac{\ker(\partial : C_k^\infty \rightarrow C_{k-1}^\infty(M))}{\text{im}(\partial : C_{k+1}^\infty(M) \rightarrow C_k^\infty(M))}, \quad (45)$$

and gives an isomorphism into the singular homology groups on  $M$ , such that the map  $\iota_* : H_k^\infty(M) \rightarrow H_k(M)$  is an isomorphism of groups.

*Remark B.6.1.* For the free abelian group of  $k$ -chains  $C_k(M)$  created by smooth  $k$ -simplices on  $M$  we introduce the notation  $C_k^\infty(M)$ . These are formal sums of finitely many smooth  $k$ -simplices. The boundary of a smooth  $(k+1)$ -simplex is thus a smooth  $k$ -chain.



*Proof.* Let  $U \subset \mathbb{R}^n$  be a subset of the Euclidean space and  $f : U \rightarrow M$  a smooth map into the manifold  $M$ . The map is said to be smooth in  $p \in U$  if there is a smooth extension of  $f$  to an open neighborhood of  $p$ . If  $p$  is arbitrary, then  $f$  is smooth in  $M$ . The standard basis vectors  $\Delta_i$  form a smooth  $k$ -simplex  $\Delta^k := [v_0, \dots, v_k] \subset \mathbb{R}^n$ , with a smooth map  $\sigma : \Delta^k \rightarrow M$ .

The inclusion  $\iota : C_k^\infty(M) \hookrightarrow C_k(M)$  satisfies  $\iota \circ \partial = \partial \circ \iota$ , from which we can see, that it induces a map  $\iota_* : H_k^\infty(M) \rightarrow H_k(M)$ , for which  $\iota_*[x] = [\iota(x)]$ .

**Lemma B.6.1.** For the simplex  $\Delta^k \subset \mathbb{R}^n$  with the standard basis vectors as vertices, let  $\sigma : \partial\Delta^k \rightarrow M$  be the continuous map with a restriction to the boundary of each face. Then  $\sigma$  is smooth as a function from  $\partial\Delta^k$  into  $M$ .

*Proof.* We follow the proof of [40, Lem. 2.1]. Let  $0 \leq i \leq k$  and  $\hat{v}_i$  denote the boundary face opposite of  $\partial_i\Delta^k = [v_0, \dots, \hat{v}_i, \dots, v_k]$ , where the primer represents the omitted element by the boundary operator. For each  $i$  and each  $v \in \partial_i\Delta^k$  we seek an open neighborhood  $v \in U_v \subseteq \mathbb{R}^n$  and a smooth extension  $\tilde{\sigma} : U_v \rightarrow M$  of  $\sigma$  such that  $\tilde{\sigma}$  restricted to  $U_v \cap \partial_i\Delta^k$  corresponds exactly to  $\sigma$ .

**Proposition B.6.1.** For each  $v \in \partial\Delta^k$  a smooth extension  $\tilde{\sigma}$  into  $M$  exists.

*Proof.* We follow the proof of [40, Lem. 2.1]. The element  $v$  is contained in at least one boundary face of  $\Delta^k$ , but cannot be contained in all of them. This means that the boundary faces  $\partial_1\Delta^k, \dots, \partial_j\Delta^k$  for  $1 \leq j \leq k$  contain  $v$ . In particular  $v$  is not contained in  $\partial_0\Delta^k$ . Let  $U = \bigcap_{i=1}^j U_i$  be the intersections for each possible  $v \in \partial_i\Delta^k$ . Using the induction hypothesis, we show that there is a smooth map  $\tilde{\sigma} : U \rightarrow M$ , whose restriction to  $U \cap \left(\bigcup_{i=1}^j \partial_i\Delta^k\right)$  matches  $\sigma$ . This property is shown locally, such that for small enough  $U$ ,  $\sigma$  maps  $U$  to a coordinate chart of  $f(v)$ , which in turn can be identified diffeomorphically as a subspace of  $\mathbb{R}^n$ . For  $j = 1$  the proof is trivial, thus assume that  $j > 1$  and that a smooth map  $\tilde{\sigma}_0 : U \rightarrow M$  exists, whose restriction to  $U \cap \left(\bigcup_{i=1}^j \partial_i\Delta^k\right)$  matches  $\sigma$ . We define  $\tilde{\sigma} : U \rightarrow M$  with

$$\tilde{\sigma} \circ (v_1, \dots, v_{i-1}, 0, v_{i+1}, \dots, v_n) = \quad (46)$$

$$= \tilde{\sigma}_0 \circ (v_1, \dots, v_m) - \tilde{\sigma}_0 \circ (v_1, \dots, v_{j-1}, 0, v_{j+1}, \dots, v_m) + \quad (47)$$

$$+ \tilde{\sigma}_j \circ (v_1, \dots, v_{j-1}, 0, v_{j+1}, \dots, v_m). \quad (48)$$

For each  $1 \leq i \leq j-1$  the restriction  $\tilde{\sigma}$  to  $U \cap \partial_i\Delta^k$  is given by

$$\tilde{\sigma} \circ (v_1, \dots, v_{i-1}, 0, v_{i+1}, \dots, v_n) = \quad (49)$$

$$= \tilde{\sigma}_0 \circ (v_1, \dots, v_{i-1}, 0, v_{i+1}, \dots, v_n) - \quad (50)$$

$$- \tilde{\sigma}_0 \circ (v_1, \dots, v_{i-1}, 0, v_{i+1}, \dots, v_{j-1}, 0, v_{j+1}, \dots, v_n) + \quad (51)$$

$$+ \tilde{\sigma}_j \circ (v_1, \dots, v_{i-1}, 0, v_{i+1}, \dots, v_{j-1}, 0, v_{j+1}, \dots, v_n) = \quad (52)$$

$$= \sigma \circ (v_1, \dots, v_{i-1}, 0, v_{i+1}, \dots, v_n). \quad (53)$$

The singular simplex  $\tilde{\sigma}_0$  is equal to  $\sigma_i$  for a  $v_i \in \Delta^k$  with  $v_i = 0$ . Similarly,  $\tilde{\sigma}_j$  is equal to  $\sigma_j$  for a  $v_j \in \Delta^k$  with  $v_j = 0$ .

An analogous argument shows, that for some face of a singular simplex  $\tilde{\sigma}$ , restricted to  $U \cap \partial_j \Delta^k$ , the inductive step can be completed as follows:

$$\tilde{\sigma} \circ (v_1, \dots, v_{j-1}, 0, v_{j+1}, \dots, v_n) = \quad (54)$$

$$= \tilde{\sigma}_0 \circ (v_1, \dots, x_{j-1}, 0, v_{j+1}, \dots, v_n) - \quad (55)$$

$$- \tilde{\sigma}_0 \circ (v_1, \dots, v_{j-1}, 0, v_{j+1}, \dots, v_n) + \quad (56)$$

$$+ \tilde{\sigma}_j \circ (v_1, \dots, v_{j-1}, 0, v_{j+1}, \dots, v_n) = \quad (57)$$

$$= \sigma \circ (v_1, \dots, v_{i-1}, 0, v_{i+1}, \dots, v_n). \quad (58)$$

□

Prop. B.6.1 proofs the lemma, as every  $v_i \in \partial \Delta^k$  has a smooth extension for arbitrary  $k$ , every element of any simplex has a smooth extension. □

For each simplex we construct a homotopy to a smooth simplex, such that the homotopies respect the restriction to the boundary faces.

**Lemma B.6.2.** For every  $k$ -simplex  $\sigma : \Delta^k \rightarrow M$  from the standard  $k$ -simplex of  $\mathbb{R}^n$  to  $M$ , there exists a continous map  $h_\sigma : \Delta^k \times I \rightarrow M$ , beeing a homotopy, satisfying the following conditions:

- (1)  $h_\sigma$  maps from  $\sigma(v) = h_\sigma(v, 0)$  to a smooth  $k$ -simplex  $\tilde{\sigma}(v) = h_\sigma(v, 1)$ .
- (2) For each boundary face inclusion  $s_{i,k} : \Delta^{k-1} \simeq \partial_i \Delta^k \hookrightarrow \Delta^k$ , we get  
 $h_{\sigma \circ s_{i,k}}(v, t) = h_\sigma(s_{i,k}(v), t)$  for  $(v, t) \in \Delta^{k-1} \times I$ .
- (3) For a smooth  $k$ -simplex  $\sigma$ ,  $h_\sigma$  is the constant homotopy, i.e.  $h_\sigma(v, t) = \sigma(v)$ .

*Proof.* For a proof we refer to [40, Lem. 2.3]. □

To finalize our proof of Theo. B.6, we follow [40, Theo. 2.4]. Let  $\iota_0 : \Delta^k \hookrightarrow \Delta^k \times I$ ,  $v \mapsto (v, 0)$  and  $\iota_1 : \Delta^k \hookrightarrow \Delta^k \times I$ ,  $v \mapsto (v, 1)$ . The smooth maps allow to define a homomorphism  $f : C_k(M) \rightarrow C_k^\infty(M)$  by mapping  $\sigma \mapsto h_\sigma \circ \iota_1$  and extending linearly. Then  $f$  is a map of chains, due to

$$f \circ \partial(\sigma) = f \left( \sum_{i=0}^k (-1)^i h_{\sigma \circ s_{i,k}} \circ \iota_1 \right) = \quad (59)$$

$$= \sum_{i=0}^k (-1)^i h_\sigma \circ (s_{i,k} \times \text{Id}_I) \circ \iota_1 = \quad (60)$$

$$= \sum_{i=0}^k (-1)^i h_\sigma \circ \iota_1 \circ s_{i,k} = \quad (61)$$

$$= \partial(h_\sigma \circ \iota_1). \quad (62)$$

Thus  $f$  induces a homomorphism  $f_* : H_k(M) \rightarrow H_k^\infty(M)$ .

**Proposition B.6.2.** The homomorphism  $f_* : H_k(M) \rightarrow H_k^\infty(M)$  is the inverse of  $\iota_* : H_k^\infty(M) \rightarrow H_k(M)$ , such that  $\iota_* = f_*^{-1}$ .

*Proof.* By Lem. B.6.2 (3)  $f \circ \iota = \text{Id}_{C_k^\infty(M)}$ , such that  $f_* \circ \iota_* = \text{Id}_{H_k^\infty(M)}$ . Vice versa, it suffices to show that a chain homotopy  $h : C_k(M) \rightarrow C_{k+1}(M)$  exists, such that  $\partial \circ h + h \circ \partial = \iota \circ f - \text{Id}_{C_k(M)}$ . Let  $V \subset \mathbb{R}^n$  and  $A : \Delta^k \rightarrow V$ , an affine transformation which sends  $v_i \mapsto p_i$  denoted as  $A(v_0, \dots, v_k)$ . We define a map  $t_{i,k} : \Delta^k \rightarrow \Delta^k \times I$  such that

$$t_{i,k} := A((v_0, 0), \dots, (v_{i-1}, 0), (v_i, 0), \dots, (v_k, 0)). \quad (63)$$

□

We use the proof from [40, Theo. 2.4] for the final step. The map  $t_{i,k} : \Delta^{k+1} \rightarrow \Delta^k \times I$  is smooth and well defined. The compositions  $t_{i,k} \circ s_{i,k+1}$  and  $t_{i-1,k} \circ s_{i,k+1}$  coincide and map from  $\Delta^k$  into  $\Delta^k \times I$ . This is stated as

$$t_{i,k} \circ s_{i,k+1} = t_{i-1,k} \circ s_{i,k+1} = \quad (64)$$

$$= A((v_0, 0), \dots, (v_{i-1}, 0), (v_i, 1), \dots, (v_k, 1)). \quad (65)$$

Analogous to this, the map  $(s_{i,k} \times \text{Id}_I) \circ t_{j,k-1} : \Delta^k \rightarrow \Delta^k \times I$  is given by

$$(s_{i,k} \times \text{Id}_I) \circ t_{j,k-1} = \begin{cases} t_{j+1,k} \circ s_{i,k+1} & \text{if } i \leq j, \\ t_{j,k} \circ s_{i+1,k+1} & \text{if } i > j. \end{cases} \quad (66)$$

We define the inverse boundary operator  $\partial^{-1} : C_k(M) \rightarrow C_{k+1}(M)$  into the  $(k+1)$ -st singular chain group:

$$\partial^{-1}(\sigma) := \sum_{i=0}^k (-1)^i h_\sigma \circ t_{i,k}. \quad (67)$$

According to [40, Lem. 2.3], applying the inverse boundary operator to the ordinary boundary operator yields

$$\partial^{-1} \circ \partial(\sigma) = \partial^{-1} \left( \sum_{i=0}^k (-1)^i \sigma \circ s_{i,k} \right) = \quad (68)$$

$$= \sum_{i=0}^k \sum_{j=0}^{k-1} (-1)^{i+j} h_\sigma \circ s_{i,k} \circ t_{j,k-1} = \quad (69)$$

$$= \sum_{i=0}^k \sum_{j=0}^{k-1} (-1)^{i+j} h_\sigma \circ (s_{i,k} \times \text{Id}_I) \circ t_{j,k-1} = \quad (70)$$

$$= \left( \sum_{0 \leq i \leq j \leq k-1} (-1)^{i+j} h_\sigma \circ t_{j+1,k} \circ s_{i,k+1} \right) + \quad (71)$$

$$+ \left( \sum_{0 \leq j \leq i \leq k-1} (-1)^{i+j} h_\sigma \circ t_{j,k} \circ s_{i+1,k+1} \right). \quad (72)$$

We interchange both operators and split the equations into subsums with  $i < j - 1$  for  $j \neq 0$  and  $i = j - 1$ ,  $j \neq k$  and  $i = j$  and finally  $i > j$ . Then the negative of the following first and last sums in Eq. 77 and 78 equal the negative of the expressions in the last two sums of Eq. 71 and Eq. 72 [40, Theo. 2.4]:

$$\partial \circ \partial^1(\sigma) = \partial \left( \sum_{i=0}^k (-1)^i h_\sigma \circ t_{i,k} \right) = \quad (73)$$

$$= \sum_{i=0}^k \sum_{j=0}^{k+1} (-1)^{i+j} h_\sigma \circ t_{i,k} \circ s_{j,k+1} = \quad (74)$$

$$= \left( \sum_{0 \leq i < j-1} \sum_{j=0}^{k+1} h_\sigma \circ t_{i,k} \circ s_{j,k+1} \right) - \quad (75)$$

$$- \left( \sum_{j=1}^{k+1} h_\sigma \circ t_{j-1,k} \circ s_{j,k+1} \right) + \quad (76)$$

$$+ \left( \sum_{j=1}^{k+1} h_\sigma \circ t_{j,k} \circ s_{j,k+1} \right) + \quad (77)$$

$$+ \left( \sum_{0 \leq j < i \leq k} (-1)^{i+j} h_\sigma \circ t_{i,k} \circ s_{j,k+1} \right). \quad (78)$$

Substituting the maps, by applying Eq. 64, cancels out Eq. 76 and Eq. 78. The composition of the boundary operator and its inverse counterpart simplifies to

$$\partial \circ \partial^{-1}(\sigma) = h_\sigma \circ t_{0,k} \circ s_{0,k+1} - h_\sigma \circ t_{k,k} \circ s_{k+1,k+1} = \quad (79)$$

$$= h_\sigma \circ A((v_0, 0), \dots, (v_k, 0)) - h_\sigma \circ A((v_0, 1), \dots, (v_k, 1)) = \quad (80)$$

$$= h_\sigma \circ \iota_1 - h_\sigma \circ \iota_0. \quad (81)$$

This shows, that the following composition cancels out, as needed:

$$\partial^{-1} \partial(\sigma) + \partial \circ \partial^{-1}(\sigma) = h_\sigma \circ \iota_1 - h_\sigma \circ \iota_0 = \quad (82)$$

$$= \iota \circ f(\sigma) - \sigma. \quad (83)$$

□

## C Supplementary Material for Sect. 5

We implement the persistence diagrams, persistence statistics and persistence landscapes using the GUDHI library [42]. For the neural nets we use the Tensorflow backend [27], with the Keras environment as wrapper [32]. We use Python in the version 3.7 as programming language. The computations of the neural nets are conducted on Intel Core i7 9700K processors and the homology related parts are computed on a NVIDIA Quadro P4000 graphic card.

**Invertible  $\mathcal{C}^1$ -Net** The neural net used in Sect. 5 is a partially invertible architecture. It consists of a total of five hidden layers. The first layer is a dense layer, which connects the input  $\mathbf{x}^{(0)} \mapsto \text{BN} \left( \sigma \left( \sum_{i=1}^n \mathbf{W}_{ji}^{(1)} \mathbf{x}_i^{(0)} + \mathbf{b}_i^{(1)} \right) \right)$  to a subspace of  $\mathbf{x}^{(0)} \in \mathbb{R}^n$ . The subspace is initially  $\mathbb{R}^{n-k}$  with  $(n-k) \leq n$ , according to Fig. 2. The sigmoid function is used as activation function. The weights are initialized uniformly distributed and the bias term is with zeros, respectively. The batch normalization can be expressed for a batch  $k$  at the  $l$ -th layer as

$$\text{BN}(\mathbf{x}_{(k)}^{(l)})_{\gamma_{(k)}^{(l)}, \beta_{(k)}^{(l)}} = \frac{\gamma}{\sqrt{\text{Var}[\mathbf{x}_{(k)}^{(l)}] + \epsilon}} \mathbf{x}_{(k)}^{(l)} + \left( \beta - \frac{\gamma \mathbb{E}[\mathbf{x}_{(k)}^{(l)}]}{\sqrt{\text{Var}[\mathbf{x}_{(k)}^{(l)}] + \epsilon}} \right), \quad (84)$$

where  $\gamma, \beta$  and  $\epsilon$  are free parameters updated during back propagation. The very first and last dense layer is not invertible. Its purpose is a learned projection into the desired subspace of dimension  $(n-k)$  for some  $k$  and afterwards projecting into the original space in an encoding fashion, see Fig. 2.

**Non-invertible  $\xi_i$ - and  $\psi_i$ -Architecture** The invertible architecture, which is described in Sect. 4.3, uses the partially linear functions  $\xi_1, \xi_2$  and  $\psi_1, \psi_2$ , which are parameterized during the learning process. The architecture is taken from the work of Ardizzone et al. [28]. These functions are to be learned and are composed in a trivially invertible way, see Sect. 4.3. The functions themselves are implemented as neural subnets. Each function is a  $\mathcal{C}^1$ -net with a total of 3 dense layers, see Sect. 4.2. The vector  $\mathbf{u}_1^{(l)} \in \text{M}(n, \mathbb{R})$  is transformed by  $\xi_1, \xi_2, \psi_1, \psi_2 : \text{M}(n, \mathbb{R}) \rightarrow \text{M}(n, \mathbb{R})$ , such that the dimension of the vector does not change. The leaky rectifying linear unit activation function has been selected for every layer in the subnet:

$$\text{LeakyReLU}(\mathbf{x}^{(l)}) = \begin{cases} \mathbf{x}_i^{(l)}, & \text{if } \mathbf{x}_i^{(l)} \geq 0, \\ 5.5 \cdot \mathbf{x}_i^{(l)} & \text{if } \mathbf{x}_i^{(l)} < 0. \end{cases} \quad (85)$$

We apply batch normalization after the activation function, see Appx. C. The partition of the vector is still an open research question [28]. We divide the flattened vector precisely into two halves, since the samples are square matrices:

$$\mathbf{x}^{(l)} = (\mathbf{u}_1^{(l)}, 0, \dots, 0) + (0, \dots, 0, \mathbf{u}_2^{(l)}). \quad (86)$$

**Hyperparameters for Learning** We use a batch size of 128 for every model. We train the neural nets for 250 epochs with the mean squared error as loss function, see Sect. 5. As optimization algorithm we use ADAM [36], with a learning rate of  $10^{-4}$ . The dataset is randomly shuffled. The high batch size should help to better recognize good parametrization. Higher batch size causes explosions of the gradient, which we visualize in Sect. 5.

## References for Appendices

27. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015), <https://www.tensorflow.org/>
28. Ardizzone, L., Kruse, J., Rother, C., Köthe, U.: Analyzing inverse problems with invertible neural networks. In: 7th International Conference on Learning Representations (2019)
29. van den Ban, E.: Lie groups. Lecture Notes in Mathematics (2003)
30. Boissonnat, J.D., Chazal, F., Yvinec, M.: Geometric and Topological Inference. Cambridge University Press (2018). <https://doi.org/10.1017/9781108297806.014>
31. Bredon, G.: Topology and geometry, vol. 139. Springer Science & Business Media (2013). <https://doi.org/10.1007/978-1-4757-6848-0>
32. Chollet, F., et al.: Keras. <https://keras.io> (2015)
33. Gruenberg, K.: The universal coefficient theorem in the cohomology of groups. Journal of the London Mathematical Society **1**(1), 239–241 (1968). <https://doi.org/10.1112/jlms/s1-43.1.239>
34. Hatcher, A.: Algebraic Topology. Cambridge University Press (2002). <https://doi.org/10.1017/S0013091503214620>
35. Hilton, P., Stammbach, U.: A course in homological algebra. Springer Science & Business Media (2012). <https://doi.org/10.1007/978-1-4684-9936-0>
36. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations (2015). <https://doi.org/10.4236/am.2011.271233>
37. Lee, J.: Introduction to smooth manifolds. Springer (2013). <https://doi.org/10.1007/978-1-4419-9982-5>
38. Munkres, J.: Elements of algebraic topology. CRC Press (2018). <https://doi.org/10.1201/9780429493911>
39. Onishchik, A., Vinberg, E., Minachin, V.: Lie groups and Lie algebras. Springer (1993). <https://doi.org/10.1007/978-3-642-57999-8>
40. Park, P.: Proof of de rham’s theorem. Harvard University (2019)
41. Pryde, A.: The five lemma for banach spaces. American Mathematical Society **65**(1), 37–43 (1977). <https://doi.org/10.1090/S0002-9939-1977-0450998-8>
42. The GUDHI Project: GUDHI User and Reference Manual. GUDHI Editorial Board, 3.1.1 edn. (2020), <https://gudhi.inria.fr/doc/3.1.1/>