Point-wise Activations and Steerable Convolutional Networks

Marco Pacini
University of Trento, Italy
Fondazione Bruno Kessler, Trento, Italy
mpacini@fbk.eu

Xiaowen Dong
University of Oxford, United Kingdom
xdong@robots.ox.ac.uk

Bruno Lepri
Fondazione Bruno Kessler, Trento, Italy
lepri@fbk.eu

Gabriele Santin
University of Venice, Italy
gabriele.santin@unive.it

Abstract

Steerable Convolutional Neural Networks are a popular and efficient class of equivariant models. For some specific groups, representations, and choice of coordinates, the most common point-wise activations, such as ReLU, are not equivariant. Hence they cannot be employed in designing equivariant neural networks. In this paper, we present a simple yet effective generalization of such results for equivariant networks. First, we prove that for groups such point-wise activations can be employed in disentangled layers only when a simple group-theoretic condition is satisfied, namely when the linear representations underlying their feature spaces are trivial representations. Second, we show analogous results for connected compact groups, where the only admitted equivariant neural networks with point-wise activations are the invariant ones.

These results demonstrate the necessity of further research for the design of suitable activation functions beyond point-wise ones.

1 Introduction

Equivariance, the property of maintaining relationships between data points through transformations, is crucial for tasks where data exhibit symmetries. Equivariant neural networks [1, 2] have shown improved generalization capabilities across various research areas, including computer vision [3–5], computer graphics [6–11], and graph learning [12–17]. Among types of equivariant neural networks, steerable Convolutional Neural Networks (CNNs) [18] have gained particular attention because of their generality. In particular, their disentangled version guarantees granular modularity of design and precise control over the number of parameters. While traditional point-wise activation functions like ReLU are commonly employed in CNNs, they lack equivariance for certain settings [6, 18]. Consequently, their application in equivariant models is restricted. In this paper, we prove that those limitations are present for a wide class of point-wise activation functions, which were introduced by Shawe-Taylor [1], and that we call *reasonable activations* (see Definition 2 for a precise definition), encompassing those commonly used in practical applications. More precisely, we show that point-wise activations can only be employed in disentangled steerable CNNs if the linear representations underlying their feature spaces are direct sum of trivial representations.

In brief, our contributions are summarized as follows: (i) we provide a complete characterization of disentangled steerable CNNs with equivariant point-wise activations, (ii) we show a stronger characterization for equivariance with respect to connected compact groups, and (iii) we discuss their impact on the most relevant cases such as Invariant Graph Networks (IGN) [12] and rotation-equivariant CNNs [19].

The paper is organized as follows: Section 2 provides an overview of related work on equivariant models and existing limitations concerning point-wise activations. Section 3 provides preliminaries for our work. In Section 4, we present the characterizations of steerable CNNs with point-wise

M. Pacini et al., Point-wise Activations and Steerable Convolutional Networks (Extended Abstract). Presented at the Second Learning on Graphs Conference (LoG 2023), Virtual Event, November 27–30, 2023.

activations and explore significant examples of equivariance with respect to the symmetric group, and the rotation group. Finally, Section 5 summarizes our findings and discusses future research directions.

2 Related Work

In recent years, equivariant neural networks have improved performance of standard neural networks by exploiting symmetries of the training data and have risen to an entirely new branch of machine learning known as geometric deep learning [20–22]. Early CNNs [23, 24] revolutionized computer vision, and by employing translation-equivariant features, they improved parameter sharing and scaling to larger datasets. Early explicit integration of representation theory, harmonic analysis, and linear invariants into machine learning can be dated to Kakarala [25] and Kondor [26]. Wood and Shawe-Taylor [1] are the first to bring equivariance into deep learning with a general approach. They define equivariant neural networks and give a classification of those models for the case of point-wise activations. In more recent years, Cohen and Welling [2, 18, 27] presented the foundational work of group equivariant convolutional networks. Precisely, they have introduced the general model of steerable CNNs [18], a popular and efficient class of equivariant models. In the following years many equivariant models with respect to symmetries up to particular transformations and different application domains appeared in the literature: rotation-invariance for galaxy morphology prediction [28], permutation invariance for set processing [7, 8], permutation invariance for graph and relational structure learning [12, 16, 29], roto-translation invariance and permutation-invariance for 3D point-cloud processing [30, 31], and roto-translation invariance for medical image analysis [32].

On the other hand, different research directions focused on theoretical aspects of equivariant models, including the creation of new frameworks [33], expressivity and universality [14, 15, 34–37], generalization bounds [33, 38–41], and characterizations [1, 22, 42]. In particular, Cohen and Welling [18] note that the equivariance of activations depends on the coordinates chosen on the representation spaces. They present some sets of admissible equivariant activations with respect to disentangled bases. Here, we prove that those sets are the *only* admissible ones. However, it should be noted that other activations beyond point-wise ones exist and are employed in practice. A good reference on this topic is Weiler and Cesa [43], and some of the presented activations are norm nonlinearities [3], squashing nonlinearities [44], tensor product nonlinearities [45], and gated nonlinearities [6].

3 Preliminaries

3.1 Representation Theory

We would like to define functions which are invariant with respect to a certain set of transformations. Interesting classes of transformations are groups [46]. A group G is a set of elements which can be composed together, can be inverted and s.t. there exist an element neutral with respect to composition. For further details we refer to Definition 3 and Example 1.

Representation theory [46] studies how abstract groups can be translated to sets of matrices which are group themselves. Given a group G, a vector space V on the field $\mathbb{R}$ of real numbers, and the set $\mathrm{GL}(V)$ of linear invertible functions from V to itself, a representation is a function $\rho : G \longrightarrow \mathrm{GL}(V)$ compatible with the group structures. When possible we will indicate such a representation by using simply V (Definition 5).

For our purposes some particular representations will play an important role. Given an action of G on a finite set X, and setting $V = \mathbb{R}^X$, a *permutation representation* is a representation such that $g(e_i) = e_{gi}$ for each $g \in G$ and $i \in X$, and a *signed permutation representation* is such that $g(e_i) = \pm e_{gi}$ (Definition 7).

In what follows we will write $\mathrm{Hom}(V, W)$ for the set of all linear maps from V to W. If V and W are vector spaces underlying representations ρ_V and ρ_W of a same group G, we define $\mathrm{Hom}_G(V, W)$ as the set of G-equivariant linear functions from V to W, i.e., the functions f compatible with representations in the sense that $f \circ \rho_V = \rho_W \circ f$. For the interests of this work, we have to consider affine maps between V and W, a composition between a linear map $\mathrm{Hom}(V, W)$ and a translation on W. We denote as $\mathrm{Aff}(V, W)$ those maps and as $\mathrm{Aff}_G(V, W)$ the set of equivariant affine functions, i.e., affine maps f such that $f \circ \rho_V = \rho_W \circ f$, see also Appendix A.3.

3.2 Steerable CNNs and Disentanglement

We follow the general framework proposed in [18] which proves the universality of the discussed model and [38] which computes generalization bounds for them. Given a group G, a G-Steerable CNN is the composition

$$\Phi = \phi_m \circ \sigma_{m-1} \circ \phi_{m-1} \circ \cdots \circ \sigma_1 \circ \phi_0 \tag{1}$$

where the V_i's are arbitrary G-representations, $\sigma_i : V_i \longrightarrow V_i$ are non-affine G-equivariant functions, and ϕ_i is an affine G-equivariant map, namely $\phi_i \in \mathrm{Aff}_G(V_i, V_{i+1})$ in our notation.

An activation $\sigma : V_i \longrightarrow V_i$ is *point-wise* if there exists a basis $\mathcal{B}_i = \{v_1, \dots, v_m\}$ of V_i and a function $\sigma' : \mathbb{R} \longrightarrow \mathbb{R}$ such that

$$\sigma(a_1 v_1 + \cdots + a_m v_m) = \sigma'(a_1) v_1 + \cdots + \sigma'(a_m) v_m$$

for each $a_1, \dots, a_m \in \mathbb{R}$.

An irreducible representation of a G-representation V_i is a minimal non-trivial G-invariant subspace and each representation V_i can be decomposed into a direct sum of irreducible spaces (Definition 9 and Theorem 3). In addition, irreducible representations are necessary for the following definition of *disentanglement* which we state in the manner of Cohen and Welling [18].

Definition 1. *A steerable CNN with point-wise activations is* ***disentangled*** *if the $\mathcal{B}_i$'s only contain vectors belonging to factors of a irreducible decomposition of V_i defined by a fixed basis.*

4 Characterization and its Implications

To clearly state the main result of the paper, we need to recall the notion of *reasonable* activation [1].

Definition 2. *A point-wise activation $\sigma : V \longrightarrow V$ is called* ***reasonable*** *if it is induced by a function $\sigma' : \mathbb{R} \to \mathbb{R}$ with the following property: There exists unique values $a^+, a^- \in \mathbb{R}$ such that the graph of $x \mapsto \sigma'(x) - \sigma'(0)$ intersects the graph of $x \mapsto a^- x$ in an infinite number of points $x < 0$, and it intersects the graph of $x \mapsto a^+ x$ in an infinite number of points $x > 0$.*

We are now able to present the main statement of this paper.

Theorem 1. *For an activation function $\sigma : V \longrightarrow V$ induced by non-odd reasonable functions, the network is disentangled if and only if the representation V is the direct sum of trivial representations.*

The proof on this theorem heavily leans on a fundamental result first appearing as Theorem 2.4 in the pioneering work of Wood and Shawe-Taylor [1]. Here, we present and prove a stronger version of it (Appendix A.5).

Theorem 2. *Let G be a compact group, $\sigma : V \longrightarrow V$ be a reasonable and equivariant activation function on a continuous G-representation V defined on the basis $\mathcal{B}$. We have the following cases.*

- *If σ is induced by an odd function then V is a signed-permutation representation.*
- *If σ is induced by a semilinear function, i.e., linear on both the positive and negative part of $\mathbb{R}$, then V is a permutation representation with respect to a positive scaling of vectors in $\mathcal{B}$.*
- *Otherwise, V is a permutation representation of G with respect to $\mathcal{B}$.*

We are now able to prove Theorem 1.

Proof. Proving the equivariance of the activation in the case V is the direct sum of trivial representations is straightforward. For the other case, thanks to Theorem 2, we can consider V to be a permutation representation. By disentanglement, we can suppose V to be irreducible and with basis $\mathcal{B} = \{v_1, \dots, v_m\}$ defining σ. Given $v \in V$, the subspace $\langle \sum_{g \in G} gv \rangle$ is one-dimensional and G-invariant. Hence, $V = \langle \sum_{g \in G} gv \rangle$ which is trivial. □

We have presented our theoretical findings on the role of equivariance in deep learning, specifically focusing on steerable CNNs and the conditions under which point-wise activations can be effectively employed. Now we study how this characterization affects and limits the design choices of networks in common practical cases such as IGNs [12] and rotation-equivariant CNNs [19]. In what follows, we only consider non-odd, reasonable and equivariant activation functions which include all the commonly used activations.

4.1 Rotation-Equivarinat Networks and Equivariance for Connected Compact Groups

We now discuss how Theorem 1 affects the design of rotation equivariant networks. The group of rotations around the origin of $\mathbb{R}^n$ is denoted as $\mathrm{SO}(3)$, it can be described as the group of real orthogonal 3×3 matrices with positive determinant (Example 1). As $\mathrm{SO}(3)$ is a compact and simply connected group [46], we will study the general case of equivariance with respect to groups presenting these two properties. We saw that each admissible continuous representation of G is the composition of an homomorphism $G \longrightarrow S_k$ for a certain k and the defining representation of S_k. As G is connected, its image in S_k is the identity element, hence the only admissible continuous G-representation is the trivial one. From now on, we will consider all the representations of compact groups to be continuous. Note that a simple input representation for rotation-equivariant networks is a vector in $\mathbb{R}^3$ which is an irreducible representation of $\mathrm{SO}(3)$ with respect to left multiplication. But, as discussed above, the only admitted output representation is the trivial one. By Schur's Lemma (Lemma 1) an equivariant map between two distinct irreducible representations is trivial. Hence, an equivariant linear layer between such spaces would collapse the entire input on zero. This means that rotation-equivariant neural network with point-wise activations for common practical scenarios have no trainable parameters and, even worse, they map all possible inputs to a single value.

4.2 On Invariant Graph Networks and Geometric Graphs

A particular class of neural networks equivariant with respect to the symmetric group are IGNs introduced by Maron et al [12] and are intimately related to graph neural networks [13, 15]. Unordered data such as sets, graphs and hypergraphs, are encoded as tensor spaces or symmetric tensor spaces which contain higher-order irreducible components. The dimension of this other components is $O(n^k)$ with k assuming all values between 0 and the order of the relation structure [48]. In the case of graphs we have $p = 2$, then there are components of dimension $O(n^2)$ which retains the widest part the input information. Because of Schur's Lemma (Lemma 1), linear layers between standard input spaces and layer representations, which are trivial by Theorem 1, destroy the majority of the input information, in the same way as described in the previous paragraph. As a simple example, in the case of sets we can consider a domain in $\mathbb{R}^n$, whose irreducible components are the trivial representation and its complement of dimension $n - 1$. Due to Schur's Lemma the complement maps to zero hence almost all the information in the input is destroyed.

Let us study the case in which two compact groups H and F act at the same time and their actions commute with each other, i.e. $G = H \times F$. We know that the irreducible representations of G are the tensor products of irreducible representations of H and F (see Appendix A.4). Hence, as $1 = \dim V \otimes W = \dim V \cdot \dim W$ if and only if $\dim V = \dim W = 1$, the only admissible disentangled components for $H \times F$ are a tensor product of admissible disentangled components for both H and F. In practice, point-clouds of k points are encoded as elements of $\mathbb{R}^3 \otimes \mathbb{R}^k$ with $SO(3) \times S_k$ acting on it [47]. Therefore, the only admissible disentangled components for point-cloud processing are the tensor products of trivial and sign representations of S_k and the trivial representation of $SO(3)$. More in general, geometric graphs or geometric p-order structures on k nodes can be encoded in $\mathbb{R}^3 \otimes (\mathbb{R}^k)^{\otimes p}$ where S_k acts as the tensor representation of p copies of the standard permutation representation $\mathbb{R}^k$ of S_k [12].

5 Conclusions and Future Directions

In conclusion, we have provided a complete characterization of disentangled steerable CNNs featuring point-wise equivariant activations. Our analysis investigates relevant examples, including IGNs and rotation-equivariant CNNs. On one hand, subsequent work will focus on analyzing the case of non-connected compact groups such as $O(3)$, the group of origin-preserving isometries of $\mathbb{R}^3$. On the other hand, we will focus on better classifying corner cases such as odd activations and signed permutation representations, or generalizing to continuous activations possibly leading to the discovery and testing of new equivariant models. Relevant future directions are the study for equivalent results for non-compact groups [6], such as isometries of the Euclidean space, and non-point-wise activations, such as norm activations [3], squashing activations [44], tensor product activations [45], and gated activations [6], as the first two activation types present slow convergence in training, the third may present high computational complexity scaling with the representation tensor power, the last is an amalgamation of the previous solutions, alleviating convergence time but not the computational burden.

References

[1] Jeffrey Wood and John Shawe-Taylor. Representation theory and invariant neural networks. *Discrete Applied Mathematics*, 69(1-2):33–60, August 1996. 1, 2, 3, 10

[2] Taco Cohen and Max Welling. Group Equivariant Convolutional Networks. In *Proceedings of The 33rd International Conference on Machine Learning*, pages 2990–2999. PMLR, June 2016. 1, 2

[3] Daniel E. Worrall, Stephan J. Garbin, Daniyar Turmukhambetov, and Gabriel J. Brostow. Harmonic Networks: Deep Translation and Rotation Equivariance. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7168–7177, July 2017. 1, 2, 4

[4] Emiel Hoogeboom, Jorn W. T. Peters, Taco S. Cohen, and Max Welling. HexaConv. In *International Conference on Learning Representations*, February 2018.

[5] Sander Dieleman, Jeffrey De Fauw, and Koray Kavukcuoglu. Exploiting Cyclic Symmetry in Convolutional Neural Networks. In *Proceedings of The 33rd International Conference on Machine Learning*, pages 1889–1898. PMLR, June 2016. 1

[6] Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco Cohen. 3D steerable CNNs: learning rotationally equivariant features in volumetric data. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, pages 10402–10413, Red Hook, NY, USA, 2018. Curran Associates Inc. 1, 2, 4

[7] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep Sets. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. 2

[8] Charles R. Qi, Su, Hao, Mo, Kaichun, and Guibas, Leonidas J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 77–85, Honolulu, HI, July 2017. IEEE. 2

[9] Haggai Maron, Or Litany, Gal Chechik, and Ethan Fetaya. On Learning Sets of Symmetric Elements. In *Proceedings of the 37th International Conference on Machine Learning*, pages 6734–6744. PMLR, November 2020.

[10] Diego Marcos, Michele Volpi, Nikos Komodakis, and Devis Tuia. Rotation Equivariant Vector Field Networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 5058–5067, Venice, October 2017. IEEE.

[11] Laurent Sifre and Stéphane Mallat. Rotation, Scaling and Deformation Invariant Scattering for Texture Discrimination. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1233–1240, June 2013. 1

[12] Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Invariant and Equivariant Graph Networks. In *International Conference on Learning Representations*, September 2018. 1, 2, 3, 4

[13] Haggai Maron, Heli Ben-Hamu, Hadar Serviansky, and Yaron Lipman. Provably Powerful Graph Networks. *International Conference of Learning Representations*, 2019. 4

[14] Haggai Maron, Ethan Fetaya, Nimrod Segol, and Yaron Lipman. On the Universality of Invariant Networks. In *Proceedings of the 36th International Conference on Machine Learning*, pages 4363–4371. PMLR, May 2019. 2

[15] Floris Geerts. On the expressive power of linear algebra on graphs, February 2020. arXiv:1812.04379 [cs]. 2, 4

[16] Risi Kondor, Truong Son Hy, Horace Pan, Brandon M. Anderson, and Shubhendu Trivedi. Covariant Compositional Networks For Learning Graphs. *International Conference of Learning Representations*, February 2018. 2

[17] Truong Hy, Shubhendu Trivedi, Horace Pan, Brandon Anderson, and Risi Kondor. Predicting molecular properties with covariant compositional networks. *The Journal of Chemical Physics*, 148:241745, June 2018. 1

[18] Taco S. Cohen and Max Welling. Steerable CNNs. In *International Conference on Learning Representations*, November 2016. 1, 2, 3

[19] Taco S. Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical CNNs. In *International Conference on Learning Representations*, February 2018. 1, 3

[20] Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges. *arXiv:2104.13478 [cs, stat]*, May 2021. arXiv: 2104.13478. 2

[21] Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric Deep Learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, July 2017. Conference Name: IEEE Signal Processing Magazine.

[22] Risi Kondor and Shubhendu Trivedi. On the Generalization of Equivariance and Convolution in Neural Networks to the Action of Compact Groups. In *Proceedings of the 35th International Conference on Machine Learning*, pages 2747–2755. PMLR, July 2018. 2

[23] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, November 1998. Conference Name: Proceedings of the IEEE. 2

[24] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, May 2017. 2

[25] Ramakrishna Kakarala. *Triple correlation on groups*. phd, University of California at Irvine, USA, 1992. UMI Order No. GAX93-04094. 2

[26] Imre Risi Kondor. *Group theoretical methods in machine learning*. Columbia University, 2008. 2

[27] Taco Cohen and Max Welling. Learning the Irreducible Representations of Commutative Lie Groups. In *Proceedings of the 31st International Conference on Machine Learning*, pages 1755–1763. PMLR, June 2014. 2

[28] Sander Dieleman, Kyle W. Willett, and Joni Dambre. Rotation-invariant convolutional neural networks for galaxy morphology prediction. *Monthly Notices of the Royal Astronomical Society*, 450(2):1441–1459, June 2015. 2

[29] Horace Pan and Risi Kondor. Permutation Equivariant Layers for Higher Order Interactions. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, pages 5987–6001. PMLR, May 2022. 2

[30] Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds, May 2018. arXiv:1802.08219 [cs]. 2

[31] Nadav Dym and Haggai Maron. On the Universality of Rotation Equivariant Point Cloud Networks. In *International Conference on Learning Representations*, October 2020. 2

[32] Erik J. Bekkers, Maxime W. Lafarge, Mitko Veta, Koen AJ Eppenhof, Josien PW Pluim, and Remco Duits. Roto-Translation Covariant Convolutional Networks for Medical Image Analysis. *Roto-Translation Covariant Convolutional Networks for Medical Image Analysis*. 1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands 2

[33] Taco S Cohen, Mario Geiger, and Maurice Weiler. A General Theory of Equivariant CNNs on Homogeneous Spaces. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. 2

[34] Siamak Ravanbakhsh. Universal Equivariant Multilayer Perceptrons. In *Proceedings of the 37th International Conference on Machine Learning*, pages 7996–8006. PMLR, November 2020. 2

[35] Ding-Xuan Zhou. Universality of deep convolutional neural networks. *Applied and Computational Harmonic Analysis*, 48(2):787–794, March 2020.

[36] Dmitry Yarotsky. Universal approximations of invariant maps by neural networks, April 2018. arXiv:1804.10306 [cs].

[37] Chaitanya K. Joshi, Cristian Bodnar, Simon V. Mathis, Taco Cohen, and Pietro Lio. On the Expressive Power of Geometric Graph Neural Networks. *International Conference of Learning Representations*, 2023. 2

[38] Arash Behboodi, Gabriele Cesa, and Taco Cohen. A PAC-Bayesian Generalization Bound for Equivariant Networks. In *Advances in Neural Information Processing Systems*, 2022. 2, 3

[39] Akiyoshi Sannai, Masaaki Imaizumi, and Makoto Kawano. Improved generalization bounds of group invariant / equivariant deep networks via quotient feature spaces. In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, pages 771–780. PMLR, December 2021.

[40] Aaron Zweig and Joan Bruna. A Functional Perspective on Learning Symmetric Functions with Neural Networks. In *Proceedings of the 38th International Conference on Machine Learning*, pages 13023–13032. PMLR, July 2021.

[41] Bryn Elesedy. Group Symmetry in PAC Learning. In *ICLR 2022 Workshop on Geometrical and Topological Representation Learning*, 2022. 2

[42] Leon Lang and Maurice Weiler. A Wigner-Eckart Theorem for Group Equivariant Convolution Kernels. In *International Conference on Learning Representations*, 2021. 2

[43] Maurice Weiler and Gabriele Cesa. General E(2) - Equivariant Steerable CNNs. *Advances in Neural Information Processing Systems*, 2019. 2

[44] Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. Dynamic Routing Between Capsules. *Advances in Neural Information Processing Systems*, 2017. 2, 4

[45] Risi Kondor. N-body Networks: a Covariant Hierarchical Neural Network Architecture for Learning Atomic Potentials, March 2018. arXiv:1803.01588 [cs]. 2, 4

[46] William Fulton and Joe Harris. *Representation Theory*, volume 129 of *Graduate Texts in Mathematics*. Springer, New York, NY, 2004. 2, 4

[47] Ben Finkelshtein, Chaim Baskin, Haggai Maron, and Nadav Dym. A Simple and Universal Rotation Equivariant Point-Cloud Network. In *Proceedings of Topological, Algebraic, and Geometric Learning Workshops 2022*, pages 107–115. PMLR, November 2022. 4

[48] Bruce E. Sagan. *The Symmetric Group*, volume 203 of *Graduate Texts in Mathematics*. Springer New York, New York, NY, 2001. 4, 10

[49] Peter Flor. On groups of non-negative matrices. *Compositio Mathematica*, 21(4):376–382, 1969. 10

A Appendix

A.1 Group Theory

We now introduce the basic concepts of group theory, which is fundamental to formalize the concept of symmetry.

Definition 3. *A* ***group*** *is a pair* $(G, \cdot)$ *where* G *is a set and* $\cdot : G \times G \longrightarrow G$ *is a function satisfying the following axioms.*

- ***Associativity:*** *for each* $g, h, k \in G$ *we have* $(g \cdot h) \cdot k = g \cdot (h \cdot k)$.
- ***Identity:*** *there exists an element* $e \in G$ *such that* $g \cdot e = e \cdot g = g$ *for each* $g \in G$.
- ***Inverse Element:*** *for each element* $g \in G$*, there exists an element* $g^{-1} \in G$ *such that* $g \cdot g^{-1} = g^{-1} \cdot g = e$.

A group is ***finite*** *if it contains a finite number of elements. A group is* ***abelian*** *or* ***commutative*** *if* $gh = hg$ *for each* $g, h \in G$.

Example 1. *Here we present some fundamental examples of groups.*

- *The set of integers with addition.*
- *The set* $\mathbb{S}^1 = \{\rho_\alpha\}$ *of rotations of angle* α *centered in the origin of the 2D Cartesian plane with composition.*
- *The set* $\mathrm{GL}(V)$ *of bijective linear maps of a vector space* V *into itself with composition. Then, the set* $\mathrm{GL}_n(\mathbb{R})$ *of* $n \times n$ *real invertible matrices form a group where the operation is row-column multiplication.*
- *Let* $\mathrm{SO}_n(\mathbb{R})$*, or simply* $\mathrm{SO}(n)$*, be the group of orthogonal matrices with positive determinant.*

- *Fix* $[n] = \{1, \dots, n\}$. *The set*

$$S_n = \{f : [n] \longrightarrow [n] \mid f \text{ is bijective}\}$$

with the composition operation form the ***symmetric group*** *or the* ***permutation group****.*

- *Given two groups* G *and* H*, the direct product* $G \times H$ *of them is still a group. The set of the elements is the cartesian product of* G *and* H *while the sum is defined as*

$$(g_1, h_1) \circ_{G\times H} (g_2, h_2) = (g_1 \circ_G g_2, h_1 \circ_H h_2).$$

Now, we introduce notion of group homomorphism, a transformation between groups which preserves the operation.

Definition 4. *A* ***group homomorphism*** *is a map*

$$\phi : G \longrightarrow H$$

between G *and* H *groups such that, for each* $g, h \in G$

$$\phi(g \cdot h) = \phi(g) \cdot \phi(h).$$

Example 2. *The map* $\Phi : \mathbb{S}^1 \longrightarrow \mathrm{GL}_2(\mathbb{R})$ *defined by*

$$\rho_\alpha \mapsto \begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{bmatrix} \tag{2}$$

is an homomorphism between the group of rotations of angle α *and the* 2×2 *invertible matrices.*

A.2 Representation Theory

Definition 5. *A* ***representation*** *of a group* G *in a vector space* V *is a group homomorphism*

$$\rho : G \longrightarrow \mathrm{GL}(V).$$

Definition 6. *Consider* $G = S_n$*.*

- $\dim V = 1$ *and* $\rho(g) = id$ *for each* $g \in S_n$*. This is called the* ***trivial*** *representation of* S_n*.*
- $\dim V = 1$ *and* $\rho(g) = sgn(g)id$*. This is called the* ***sign*** *representation of* S_n*.*

Definition 7. *A representation* $\rho : G \longrightarrow \mathrm{GL}_n(\mathbb{R})$ *is*

- ***Non-negative*** *if all the elements of each* $\rho(g)$ *are non-negative.*
- ***Monomial*** *if each row and column of each* $\rho(g)$ *has exactly one non-zero element.*
- *A* ***permutation*** *representation if it is monomial and each non-zero element is* 1.
- *A* ***signed permutation*** *representation if it is monomial and each non-zero element is* ± 1.

Definition 8. *A linear map* $\Phi : V \longrightarrow W$ *is* G***-equivariant*** *with respect to the representations* $\rho_1 : G \longrightarrow \mathrm{GL}(V)$ *and* $\rho_2 : G \longrightarrow \mathrm{GL}(W)$ *if*

$$\rho_2 \circ \Phi = \Phi \circ \rho_1.$$

We will denote the space of all G*-equivariant maps between* ρ_1 *and* ρ_2 *by* $Hom_G(\rho_1, \rho_2)$ *or* $Hom_G(V, W)$ *when* ρ_1 *and* ρ_2 *will be clear from the context.*

Definition 9. *A representation* $\rho : G \longrightarrow \mathrm{GL}(V)$ *is* ***irreducible*** *if there exists no non-trivial subspace* W *of* V *such that* $\rho(g)(W) \subseteq W$ *for each* $g \in G$*.*

An important result in representation theory of finite groups states that there is always a decomposition into irreducible representations.

Theorem 3. *For each representation* $\rho : G \longrightarrow \mathrm{GL}(V)$*, there exists a decomposition*

$$V = V_1 \oplus \cdots \oplus V_m$$

where each V_i *is irreducible for* ρ*. This decomposition is unique up to isomorphism and permutation of the factors.*

Another tool coming from Representation Theory is Schur's Lemma which will be fundamental to understand our results.

Lemma 1. *Let V and W be non-isomorphic irreducible representations, there is only one G-equivariant linear map between them and it is the trivial one.*

Definition 10. *The fixed set for a representation $\rho : G \longrightarrow \mathrm{GL}(V)$ is $V^G = \{v : gv = v\}$. Note that V^G is a representation for G and the action is trivial.*

Theorem 4. *The following properties are true for representations of a finite group G.*

- $\dim V^G$ *is the multiplicity of the trivial representation in V,*
- $\dim \mathrm{Hom}_G(V, W) = \dim(V \otimes W)^G$.

A.3 Affine maps

Definition 11. *Let V and W be two $\mathbb{K}$-vector spaces and define the translation of a vector w in W as a non-linear bijective map $\tau_w : v \mapsto v + w$. Define the space of **affine maps** from V to W as*

$$\mathrm{Aff}(V, W) = \{\tau_w \circ f | w \in W \text{ and } f \in \mathrm{Hom}(V, W)\}.$$

Note that is a more general definition with respect to the standard one, where f is an isomorphism of a vector space V.

Theorem 5. *The decomposition of an affine map $\phi \in \mathrm{Aff}(V, W)$ in translational part τ_w and f linear part is unique.*

Proof.

$$\tau_{w_1} \circ f_1 = \phi = \tau_{w_2} \circ f_2,$$

evaluating in 0 leads to

$$w_1 = \phi(0) = w_2.$$

Write $w = w_1 = w_2$, and note that

$$\tau_w \circ f_1 = \tau_w \circ f_2,$$

by the bijectivity of translations,

$$f_1 = f_2.$$

□

Let V and W be G-representation. An affine map $\phi \in \mathrm{Aff}(V, W)$ is G-equivariant if $\phi \circ g = g \circ \phi$ for each $g \in G$, write the set of G-equivariant affine maps from V to W as $Aff_G(V, W)$.

Theorem 6. *$\phi = \tau_w \circ f \in \mathrm{Aff}_G(V, W)$ if and if $f \in \mathrm{Hom}_G(V, W)$ and v is invariant.*

Proof. Note that for each $g \in G$,

$$g \circ \tau_w \circ f = \tau_{gw} \circ (g \circ f).$$

Observe that

$$\phi \circ g = g \circ \phi,$$

if and only if

$$\tau_w \circ f \circ g = g \circ \tau_w \circ f = \tau_{gw} \circ (g \circ f)$$

if and only if, by the previous proposition,

$$w = gw$$

and

$$f \circ g = g \circ f$$

for each $g \in G$. □

A.4 Representations of Group Products

Remark 1. *Let $V \otimes W$ be a finite-dimensional $G \times H$-representations and V_i's a complete list of irreducible G-representations and W_j's a complete list of irreducible H-representations, then $V_i \otimes W_j$'s is a complete list of irreducible $(G \times H)$-representations (See [48] for proofs in case G and H are finite). If m_i is the multiplicity of V_i in V and n_j is the multiplicity of W_j in W then the multiplicity of $V_i \otimes W_j$ in $V \otimes W$ is $m_i n_j$. This can be easily seen by writing the irreducible decompositions of V and W and use the distributive property of tensor products and direct sums. Note that the same is true if G and H are compact groups and the representations are continuous. If S is an G-isotypic component of $V \times W$ of type ρ then it is H-invariant and each $h \in H$ acts as G-equivariant endomorphism of S. By Lemma 1, $S = \rho \otimes \sigma$, which is $\bigoplus_i \rho \otimes \sigma_i$, where σ_i are irreducible H-representations. Iterating the decomposition if necessary and by the finite dimension of V and W, we conclude.*

A.5 Characterization Theorem

The characterization theorem presented by Wood and Shawe-Taylor [1] differs from the statement of Theorem 2 which is actually stronger. The original statement by Wood and Shawe-Taylor for reasonable non-affine activations can be stated as follows:

Theorem 7. *Let $\sigma : V \longrightarrow V$ be reasonable and equivariant activation function on a representation V defined on the base $\mathcal{B}$ of a G-representation of a finite group G. We have the following cases.*

- *If σ is induced by an odd function then V is a signed-permutation representation.*
- *If σ is induced by a non-odd semilinear function, i.e., linear on both the positive and negative part of $\mathbb{R}$, then V is a non-negative monomial representation with respect to $\mathcal{B}$.*
- *Otherwise, V is a permutation representation of G with respect to $\mathcal{B}$.*

We now present the proof of Theorem 2 using Theorem 7.

Proof. Looking at the proof of Theorem 7, we see that it works seamlessly for the general case of continuous representations of compact groups. Additionally, note that an odd semilinear function is linear, as activations are non-affine, we can just consider semilinear functions in the second element of the bullet list. Finally, leveraging the work by Flor [49] on groups of non-negative matrices, non-negative monomial representations are isomorphic to permutation representations through a positive scaling of basis. In this way, the second element in the bullet list of Theorem 7 reduces to the second element of Theorem 2. □