Towards a Theory of Neural Topology: The Homology of Rectified Units

William H. Guss [1] **Ruslan Salakhutdinov** [1]

The success of deep learning in vision (Krizhevsky et al., 2012), natural language processing (Wu et al., 2016), and reinforcement learning (Mnih et al., 2015) has been accompanied by a troubling gap between theory and practice: the power of deep models to learn subject to geometric constraints on data is not fully understood.

This central question of the expressivity of different neural architectures has been the subject of much analysis: stemming from the universal approximation results of Cybenko (1989); Hornik (1991), several authors have studied expressivity in the lense of VC theory (Anthony & Bartlett, 2009), boolean circuit theory (Legenstein & Maass, 2001), and reproducing kernel hilbert spaces (Eldan & Shamir, 2016). From a primarily geometric perspective several papers have analyed the benefits of depth in the lense of Riemannian geometry (Poole et al., 2016), the theory of hyperplane arrangements (Raghu et al., 2016), and algebraic topology (Guss & Salakhutdinov, 2018; Bianchini & Scarselli, 2014).

Among the foregoing, topological expressivity theory is powerful in that it not only characterizes network power but also yields computable procedures for architecture selection. The work of Bianchini & Scarselli (2014) studies the homology of sublevel sets of the decision boundaries of neural networks, and effectively uses homology to show that deeper models are exponentially more expressive in the homology of their decision boundaries than their shallow counterparts. Computationally, Guss & Salakhutdinov (2018) gave an empirical characterization of how the persistent homology of the data relates to the efficacy of learning and expressive power of neural networks directly. These authors showed that there by understanding the relationship between topological invariants of data and the *decision regions* of models, expressivity results of a topological flavor have great practical application in choosing architectures. However, an exact topological characterization remains unsolved.

In this work we show that the topology of neural networks is deeply connected to the algebra of hyperplane arrangements; that is beyond the number of linear regions, hyperplane arrangement theory provides a prescriptive procedure for computing the exact homology of neural decision boundaries directly from their weights. The contributions of our theoretical framework are threefold:

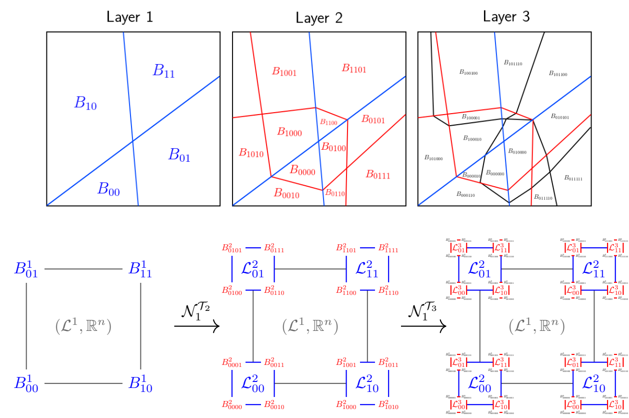

Figure 1. The HHIL for $N_3^{\mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3}$, a two hidden layer neural network with two neurons on every hidden layer.

1. *A decomposition of the decision regions of multilayer neural networks into a closed cover of convex polytopes can be given directly as an expression of the weights.* Extending the results of Raghu et al. (2016), we give a closed form formula for the output of neural networks on regions of the type in Figure 1 labeled by bit strings.
2. *The polytope cover corresponds to the leaf nodes of a hierarchical semilattice (HHIL) induced by the intersections of the halfspace activations regions of neural units.* This correspondence depicted above is shown to be order isomorphic to a rotation of the hyperplane arrangement lattice of the network.
3. *The topology of deep decision regions is equivalent to the combinatorics of their HHILs in a natural way*; that is, the nerve construction of the convex polytope cover has the same homology as that of the lattice viewed as a topological space (modulo a linear sublattice).

These results are a first step in developing a general theory of neural topology, and it is the subject of future work to study how the structure and representation of various algebraic structures on HHILs yield new methods for comparing the power of neural architectures.

[1]Machine Learning Department, Carnegie Mellon University, Pittsburgh PA.

References

Anthony, Martin and Bartlett, Peter L. *Neural network learning: Theoretical foundations*. cambridge university press, 2009.

Bianchini, Monica and Scarselli, Franco. On the complexity of neural network classifiers: A comparison between shallow and deep architectures. *IEEE transactions on neural networks and learning systems*, 25(8):1553–1565, 2014.

Cybenko, George. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

Eldan, Ronen and Shamir, Ohad. The power of depth for feedforward neural networks. In *Conference on Learning Theory*, pp. 907–940, 2016.

Guss and Salakhutdinov. On characterizing the capacity of neural networks using algebraic topology. 2018. URL https://openreview.net/forum?id=H11lAfbCW.

Hornik, Kurt. Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257, 1991.

Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012.

Legenstein, Robert A and Maass, Wolfgang. Foundations for a circuit complexity theory of sensory processing. In *Advances in neural information processing systems*, pp. 259–265, 2001.

Mnih, Volodymyr, Kavukcuoglu, Koray, Silver, David, Rusu, Andrei A, Veness, Joel, Bellemare, Marc G, Graves, Alex, Riedmiller, Martin, Fidjeland, Andreas K, Ostrovski, Georg, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529, 2015.

Poole, Ben, Lahiri, Subhaneil, Raghu, Maithra, Sohl-Dickstein, Jascha, and Ganguli, Surya. Exponential expressivity in deep neural networks through transient chaos. In *Advances in neural information processing systems*, pp. 3360–3368, 2016.

Raghu, Maithra, Poole, Ben, Kleinberg, Jon, Ganguli, Surya, and Sohl-Dickstein, Jascha. On the expressive power of deep neural networks. *arXiv preprint arXiv:1606.05336*, 2016.

Wu, Yonghui, Schuster, Mike, Chen, Zhifeng, Le, Quoc V, Norouzi, Mohammad, Macherey, Wolfgang, Krikun, Maxim, Cao, Yuan, Gao, Qin, Macherey, Klaus, et al. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*, 2016.