Complexity of Injectivity and Verification of ReLU Neural Networks

Vincent Froese VINCENT.FROESE@TU-BERLIN.DE
Technische Universität Berlin, Faculty IV
Institute of Software Engineering and Theoretical Computer Science, Algorithmics and Computational Complexity.

Moritz Grillo GRILLO@MATH.TU-BERLIN.DE
Martin Skutella MARTIN.SKUTELLA@TU-BERLIN.DE
Technische Universität Berlin, Faculty II, Institute of Mathematics,
Combinatorial Optimization and Graph Algorithms.

Editors: Nika Haghtalab and Ankur Moitra

Neural networks with ReLU activation play a key role in modern machine learning. Understanding the functions represented by ReLU networks is a major topic in current research as this enables a better interpretability of learning processes.

Injectivity of a function computed by a ReLU network, that is, the question whether different inputs to the network always lead to different outputs, plays a crucial role whenever invertibility of the function is required, such as, e.g., for inverse problems or generative models. The exact computational complexity of deciding injectivity was recently posed as an open problem (Puthawala et al., 2022). We resolve this by proving that deciding injectivity is coNP-complete for networks with any number of layers (that is, already for a single ReLU-layer). On the positive side, we show that for a single ReLU-layer the problem is still tractable for small input dimension; more precisely, we present a parameterized algorithm which yields fixed-parameter tractability with respect to the input dimension.

We further investigate the network verification problem, that is, deciding whether all inputs from a given domain map to outputs in a prescribed set. This is of great importance since neural networks are increasingly used in safety-critical systems. While prior hardness results focused on specific input sets (e.g., unit cubes), we strengthen these results by proving coNP-hardness for all input sets containing a ball (under mild assumptions). This establishes that the intractability is intrinsic to the ReLU network and not to special geometries of input sets. We also show that approximating the maximum of ReLU networks within any constant factor is infeasible unless P = NP.

To complement the study of injectivity, we initiate the first systematic analysis of surjectivity for ReLU networks. We characterize surjectivity for networks with one-dimensional output, derive a polynomial-time algorithm for constant input dimension, and prove NP-hardness in general. Finally, we relate the surjectivity question to zonotope containment, linking our results to problems in computational convexity and control theory.

Our hardness reductions reveal interesting connections between graph cut problems and ReLU network properties via graphical hyperplane arrangements that might be of independent interest.[1]

1. Extended abstract. Full version available at https://arxiv.org/abs/2405.19805.

References

Michael Puthawala, Konik Kothari, Matti Lassas, Ivan Dokmanic, and Maarten V. de Hoop. Globally injective ReLU networks. *Journal of Machine Learning Research*, 23:105:1–105:55, 2022.