

# Algebraic Neural Network Theory

**Kathlén Kohn**

KTH, *kathlen@kth.se*

The space of functions parametrized by a fixed neural network architecture is known as its "neuromanifold", a term coined by Amari. Training the network means to solve an optimization problem over the neuromanifold. Thus, a complete understanding of its intricate geometry would shed light on the mysteries of deep learning. This talk explores the approach to approximate neural networks by algebraic ones that have semialgebraic neuromanifolds. In this setting, we can interpret training the network as finding a "closest" point on the neuromanifold to some data point in the ambient space. In particular, this perspective enables us to see that the singularities (and boundary points) of the neuromanifold can cause a tradeoff between efficient optimization and good generalization to unseen data. We discuss several conjectures on singularities and hidden parameter symmetries (i.e., fibers) of neuromanifolds. This talk will focus on 3 popular architectures: multi-layer perceptrons, convolutional networks, and self-attention networks that are the key ingredient of large language models like ChatGPT. The results presented in this talk are based on several joint works with Paul Breiding, Erin Connelly, Nathan Henry, Giovanni Marchetti, Stefano Mereta, Vahid Shahverdi, and Matthew Trager.