Deep Barron Classes Reinhold Schneider (Kunoth & M. Oster)

We exploit the ideas of Barron spaces and the corresponding representation as continuous interpretation of infinitely wide shallow networks and neural ODEs as infinite deep residual network architectures. Barron spaces are characterizing approximation rates of shallow neural networks, formally without curse of dimensions. Barron functions are supposed to have continuous representation formula

$$h(x) = \int_U a\sigma(Ax+b)d\mu_1(a,A,b) \ ,$$

with a random measure μ_1. The discrete versions are based on approximation of the measure by Dirac delta's (e.g. in the weak-* sense)

$$\mu_1 = \frac{1}{N}\sum_{i=1}^{N} \delta_{a^i,A^i,b^i}, (a^i,A^i,b^i) \in U \ \Rightarrow h(x) = \frac{1}{N}\sum_{i=1}^{N} a^i\sigma(A^ix+b^i) \ ,$$

where the a^i, A^i, b^i are sampled from the random measure μ_1. Convergence can be deduced from error estimates of Monte Carlo quadrature. More details can be found e.g. in [2] In order to consider deep neural networks, we remind that a large family of deep Neural Networks, e.g ResNets, can be view as an Euler discretization of an ODE $\dot{x}(t) = v(t,x)$, $x(0) := x$, with a driving field v given as a shallow NN. This leads to the following feedback optimal control problem.

We seek to learn a function f by deep neural networks with activation function σ. An abstract feedback optimal control problem with measure-valued controls $\mu_2(t)$ of the form

$$\min_{\mu(\cdot)} \mathcal{J}(\mu(\cdot)), \quad \mathcal{J}(\mu(\cdot)) := \int_{\mathbb{R}^d} \|f(x) - \int a\sigma(A\ z(T,x)+b)\ d\mu_1(t;a,A,b)\|^2 dx$$

$$\text{such that} \quad \frac{d}{dt}z(t,x) = \int a\sigma(Az(t,x)+b)\, d\mu_2(t;a,A,b), \quad z(0,x) = x$$

The above control problem can be interpreted as an infinite deep neural network: Except the last layer is of another particular form. We discuss the issue of existence and the treatment of gradient schemes. We show the existence of minimizers to the optimal control problem by using Prokhorov's theorem on tight measures and some regularity assumptions on the activation function [4]. The present perspective provides an interesting relationship to HamiltonJacobi Bellmann equations and Potential Mean Field Games.

References

[1] L. Ambrosio, N. Gigli and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Birkhäuser (2005).

[2] E, Weinan and Wojtowytsch, Stephan,
Some observations on high-dimensional partial differential equations with Barron data, arXiv (2020,)

[3] Z. Ding, S. Chen, Q. Li and S.J. Wright. Overparameterization of DeepResnet: Zero Loss and Mean-field Analysis. Journal of Machine Learning Research **23.48** (2022), 1–65.

[4] M. Oster. *On Semi-Global Optimal Control Problems and their Applications in Machine Learning (Working Title)*, Doctoral Dissertation, Institute for Mathematics, Faculty Mathematics and Natural Sciences, Technical University of Berlin, submitted, 2023.