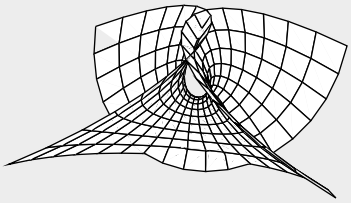

Oberseminar Mathematische Strömungsmechanik

Institut für Mathematik der Julius-Maximilians-Universität Würzburg

Dmitry Yarotsky

Skoltech, Moscow

Spectral analysis of neural network learning

Abstract:

Neural networks are complex models, and their training - usually performed by some version of gradient descent - is not easy to analyze theoretically. The problem becomes tractable for linearized networks, e.g., those trained in the NTK regime or being close to convergence. In this setting the problem reduces to optimization of an ill-conditioned quadratic problem, which can be described in terms of its spectral characteristics. We find that in many realistic network training scenarios the respective spectral distributions are well-fitted by power laws, and we show that these power laws can be derived theoretically under some reasonable assumptions. Under spectral power laws, convergence of optimization also follows a power law, with different exponents depending on the version of gradient descent. A particularly important case is the mini-batch stochastic gradient descent (SGD) with momentum. This algorithm is characterized by a rich phase diagram with two convergence phases. We derive an explicit theoretical stability condition for mini-batch SGD and demonstrate some other phenomena, e.g. that in some problems the optimal momentum parameter has a negative value.

Raum 40.03.003 (Emil Fischer Str. 40) , the speaker joins us via Zoom

Thursday, Jan. 12 at 12:30 pm

Zu diesem Vortrag sind Sie herzlich eingeladen.

gez. Christian Klingenberg