4TU.AMI MATHEMATICS FOR INNOVATION

4TU.AMI Conference
"Mathematics on Deep Learning"

Remi Gribonval - Centre de Recherche INRIA Rennes

Approximation with sparsely connected deep networks

Many of the data analysis and processing pipelines that have been carefully engineered by generations of mathematicians and practitioners can in fact be implemented as deep networks. Allowing the parameters of these networks to be automatically trained (or even randomized) allows to revisit certain classical constructions.

The talk first describes an empirical approach to approximate a given matrix by a fast linear transform through numerical optimization. The main idea is to write fast linear transforms as products of few sparse factors, and to iteratively optimize over the factors. This corresponds to training a sparsely connected, linear, deep neural network. Learning algorithms exploiting iterative hard-thresholding have been shown to perform well in practice, a striking example being their ability to somehow "reverse engineer" the fast Hadamard transform. Yet, developing a solid understanding of their conditions of success remains an open challenge.

We then proceed to study the expressivity of deep neural networks. Measuring a network's complexity by its number of connections or by its number of neurons, we consider the class of functions for which the error of best approximation with networks of a given complexity decays at a certain rate when increasing the complexity budget. Using results from classical approximation theory, we show that this class can be endowed with a (quasi)-norm that makes it a linear function space, called approximation space. We establish that allowing the networks to have certain types of "skip connections" does not change the resulting approximation spaces. We also discuss the role of the network's nonlinearity (also known as activation function) on the resulting spaces, as well as the role of depth. For the popular ReLU nonlinearity and its powers, we relate the newly constructed spaces to classical Besov spaces. The established embeddings highlight that some functions of very low Besov smoothness can nevertheless be well approximated by neural networks, if these networks are sufficiently deep.

Joint work with Luc Le Magoarou (Inria), Gitta Kutyniok (TU Berlin), Morten Nielsen (Aalborg University) and Felix Voigtlaender (KU Eichstätt).