Building Foundations for Scientific Machine Learning at Scale

Delivering on the promise of machine learning leading to transformational changes in science will require going beyond using machine learning methods (typically developed for very different industrial applications and implemented in very different computational environments) to developing novel machine learning methods that are well-suited to the needs of science. I will describe several examples of this from recent work: one that used methods from numerical integration theory to go beyond traditional machine learning training/testing methodology on discrete data to test whether the learned function is meaningfully continuous; and another that uses ideas from disordered system theory to develop metrics that can be used to evaluate the quality of trained models, even without access to training or testing data. A theme in each case is to use scientific ideas to develop machine learning methods that can be applied to science data and beyond.

In these and other cases, the core computational component of machine learning methods typically boils down to matrix or graph or optimization algorithms. Here too, the needs of scientific machine learning differ from the needs of industrial machine learning. Thus, I will also describe how we are incorporating randomized matrix algorithms within LAPACK (RandLAPACK). Several challenges and opportunities will be outlined.

Chair: Prof. Sabine Le Borne

For the Zoom link, please write a short email to dashh-office@desy.de or subscribe here (https://lists.desy.de/sympa/subscribe/datasciencecolloquium) to our mailing list.