

Deep Neural Networks motivated by PDEs

Lars Ruthotto¹

One of the most promising areas in artificial intelligence is deep learning, a form of machine learning that uses neural networks containing many hidden layers. Recent success has led to breakthroughs in applications such as speech and image recognition. However, more theoretical insight is needed to create a rigorous scientific basis for designing and training deep neural networks, increasing their scalability, and providing insight into their reasoning.

This talk bridges the gap between partial differential equations (PDEs) and neural networks and presents a new mathematical paradigm that simplifies designing, training, and analyzing deep neural networks. It shows that training deep neural networks can be cast as a dynamic optimal control problem similar to path-planning and optimal mass transport. The talk outlines how this interpretation can improve the effectiveness of deep neural networks. First, the talk introduces new types of neural networks inspired by to parabolic, hyperbolic, and reaction-diffusion PDEs. Second, the talk outlines how to accelerate training by exploiting multi-scale structures or reversibility properties of the underlying PDEs. Finally, recent advances on efficient parametrizations and derivative-free training algorithms will be presented.

The talk is joint work with the groups of Eldad Haber and Eran Treister and based in part on [1, 2, 3, 4].

References:

- [1] <https://arxiv.org/abs/1705.03341>
- [2] <https://arxiv.org/abs/1703.02009>
- [3] <https://arxiv.org/abs/1709.03698>
- [4] <https://arxiv.org/abs/1804.04272>

¹Emory University, Atlanta, Dep. of Mathematics and Computer Science
lruthotto@emory.edu