Deepest Learning Using Stochastic Partial Differential Equations

Graduate student project

within “Mathematics of AI” in WASP

AI and partial differential equations

Machine learning is an important research direction in artificial intelligence; in particular, deep learning has become very popular in recent years. Networks are trained on observations in order to be able to act independently of humans from some point on. Mathematically, the learning and training process consist of minimizing a *loss function* over a set of weights in a (deep) neural network. While a simple gradient descent approach has shown to work well for many practical problems, a rigorous mathematical justification is still missing. Another way to optimize the network is stochastic gradient descent methods, for which the first mathematical results on convergence are given in the recent preprint [2].

This project is based on a different approach. In [1], an equivalence is shown between deep learning, as a part of machine learning, and statistical data assimilation, as it is widely used in physical and biological sciences. Here, the number of layers in a deep learning network on the AI side corresponds to (discrete) time in the data assimilation framework. Therefore, the search of a global minimum of a cost functional, in order to optimize the weights in the network, can be translated to a data assimilation problem. If one goes to the limit in the sense that time or, equivalently, the layer label becomes continuous, one obtains a partial differential equation. The authors in [1] call this limit *deepest learning*, since it can be seen as the number of layers tending to infinity in the deep learning framework. The resulting Euler–Lagrange equations turn out to be a two point boundary value problem, which can be considered within the framework of variational methods. It is shown that the equations respect a symplectic symmetry and both Lagrangian and Hamiltonian versions of these problems are derived in [1]. The resulting equations are the starting point of the suggested project.

In conclusion, the connection between partial differential equations like the Euler–Lagrange equations, which have been used in statistical physics for similar problems, and deep learning is pointed out in [1], which takes into account that such minimization problems have been around for a long time. This point of view transforms the original minimization problem in the network into the study of a partial differential equation and makes partial differential equations with their analysis techniques and numerical approximation methods readily available as mathematical tools in deep learning. The next step is to transform this approach to a stochastic problem. This is part of the following project description.

Research area

The project will address existence, uniqueness, and regularity of solutions and the development of computational methods for the Euler–Lagrange differential equations derived in [1]. Known methods are likely to get stuck in local minima or saddle points and adding noise in a similar way as in [3] and [4] can lead to much improved solutions. This transforms the problem of network optimization further into a stochastic partial differential equation that has to be analyzed and solved. Similarly to the deterministic equations, questions on analytical properties of solutions and on their approximation by computational efficient and robust discretization schemes have to be answered for the stochastic counterparts. Furthermore, the application of the new methods to the original machine learning problems and the testing of their performance in practice are an important and challenging part of our project.

The proposers have the relevant background and experience of research in deterministic and stochastic partial differential equations and numerical analysis. In the suggested PhD project this knowledge can be brought to bear on machine learning.

The overall goal of the project is to develop a better mathematical understanding of deep learning, leading to reliable and efficient computational techniques, by using knowledge from stochastic partial differential equations. This can pave the way for new and better machine learning algorithms.

The PhD student will be supervised by *Annika Lang* in close collaboration with Mihály Kovács, Stig Larsson, and Klas Modin from the Department for Mathematical Sciences, Division of Applied Mathematics & Statistics at Chalmers University of Technology and the University of Gothenburg. An industry cooperation with Adam Andersson at Syntronic AB is planned. It is very important to attack the research project from both perspectives from the beginning. The interaction of the work on the theoretical, proof based questions rised above and of the first implementations will enhance the understanding of both components and lead faster to even better results.

References

[1] Henry Abarbanel, Paul Rozdeba, and Sasha Shirman. Machine learning, deepest learning: Statistical data assimilation problems. arXiv:1707.01415 [cs.AI], July 2017.

[2] Arnulf Jentzen, Benno Kuckuck, Ariel Neufeld, and Philippe von Wurstemberger. Strong error analysis for stochastic gradient descent optimization algorithms. arXiv:1801.09324 [math.NA], January 2018.

[3] Olivier Juan, Renaud Keriven, and Gheorghe Postelnicu. Stochastic Motion and the Level Set Method in Computer Vision: Stochastic Active Contours. *International Journal of Computer Vision*, 69(1):7–25, 2006.

[4] Annika Lang. *Simulation of stochastic partial differential equations and stochastic active contours*. PhD thesis, Universität Mannheim, 2007.