

## **Analyzing Optimization in Deep Learning via Trajectories**

Nadav Cohen, Institute for Advanced Study

The prominent approach for analyzing optimization in deep learning is based on the geometry of loss landscapes. While this approach has led to successful treatments of shallow (two layer) networks, it suffers from inherent limitations when facing deep (three or more layer) models. In this talk I will argue that a more refined perspective is in order, one that accounts for the specific trajectories taken by the optimizer. I will then demonstrate a manifestation of the latter approach by analyzing the trajectories of gradient descent over arbitrarily deep linear neural networks. We will derive what is, to the best of my knowledge, the most general guarantee to date for \*efficient\* convergence to global minimum of a gradient-based algorithm training a deep network. Moreover, in stark contrast to conventional wisdom, we will see that, sometimes, gradient descent can train a deep linear network faster than a classic linear model. In other words, depth can accelerate optimization, even without any gain in expressiveness, and despite introducing non-convexity to a formerly convex problem.

Works covered in this talk were in collaboration with Sanjeev Arora, Noah Golowich, Elad Hazan and Wei Hu.