CRUNCH Seminars at Brown, Division of Applied Mathematics

Friday – May 10, 2019

Paper Review: Deep relaxation: partial differential equations for optimizing deep neural networks by Pratik Chaudhari, Adam Oberman and Stanley Osher, et al.

Yeonjong Shin

Entropy-SGD is a first-order optimization method which has been used successfully to train deep neural networks. This algorithm, which was motivated by statistical physics, is now interpreted as gradient descent on a modified loss function. The modified, or relaxed, loss function is the solution of a viscous Hamilton–Jacobi partial differential equation (PDE). Experimental results on modern, high-dimensional neural networks demonstrate that the algorithm converges faster than the benchmark stochastic gradient descent (SGD). Well-established PDE regularity results allow us to analyze the geometry of the relaxed energy landscape, confirming empirical evidence. Stochastic homogenization theory allows us to better understand the convergence of the algorithm. A stochastic control interpretation is used to prove that a modified algorithm converges faster than SGD in expectation.

paper link: https://link.springer.com/article/10.1007/s40687-018-0148-y
arXiv link: https://arxiv.org/abs/1704.04932