PDE and control methods for optimizing deep neural networks

Martino Bardi

University of Padova

email: bardi@math.unipd.it

Abstract

A central problem in deep learning of neural networks is the minimization of a highly nonlinear, non-convex, possibly non-smooth function of a large number of variables. In particular one seeks robust minima of the loss function with a large basin of attraction. An entropy gradient descent method was proposed and studied in [1]. The related paper [2] developed it into an algorithm called *Deep Relaxation* that is justified by a homogenization limit in a system of stochastic differential equations.

We reformulate such limit and generalize it as a singular perturbation problem for 2nd order Hamilton-Jacobi-Bellman equations. We treat it by viscosity methods and with ideas from ergodic control reminiscent of weak KAM theory. We prove some convergence results that give a rigorous mathematical framework for the algorithm in [2] as well as for various extensions.

The talk is about some joint work with Hicham Kouhkouh, also part of his PhD thesis [3].

[1] Chaudhari, P., Choromanska, A., Soatto, S., LeCun, Y., Baldassi, C., Borgs, C., ... & Zecchina, R.: Entropy-SGD: Biasing gradient descent into wide valleys. J. Stat. Mech. 2019

[2] Chaudhari, P., Oberman, A., Osher, S., Soatto, S., & Carlier, G. Deep relaxation: partial differential equations for optimizing deep neural networks. Res. Math. Sci. 2018

[3] Kouhkouh, H.: Some asymptotic problems for Hamilton-Jacobi-Bellman equations and applications to global optimization, PhD thesis 2022, University of Padova.