Entropic Gradient Descent and relaxation in deep learning via stochastic control.

Martino Bardi

DM, Università di Padova

bardi@math.unipd.it

Global unconstrained optimization of a non-convex function f in Euclidean space is a classical problem that is attracting considerable new attention in view of applications to the training of deep neural networks, especially for highly non-linear, possibly non-smooth loss functions f in very high space dimension. Most modern algorithms are based on Sthochastic Gradient Descent, that requires a smoothing of the loss function f. Local Entropy is a recent regularization procedure which, coupled with a relaxation of the dynamical system, appears very effective in practical problems. To understand the success of this algorithm, Chaudary, Oberman, Osher, Soatto, and Carlier proposed a singular perturbation problem whose limit via homogenisation is the Entropic Gradient Descent.

We make their result precise and give a rigorous proof. Next we generalize it by treating a tuning parameter, the learning rate, as a dynamic control. We embed this in a much larger class of singular perturbations of stochastic control systems and prove the convergence of value functions and of optimal trajectories.

This is joint work with Hicham Kouhkouh (arXiv:2209.05564).