

# On the dynamics of Neural Network Optimization

JOAN BRUNA

CIMS New York University

Virtually all modern deep learning systems are trained with some form of local descent algorithm over a high-dimensional parameter space. Despite its apparent simplicity, the mathematical picture of the resulting setup contains several mysteries that combine statistics, approximation theory and optimization.

In order to make progress, authors have recently focused in the so-called ‘overparametrised’ regime, which studies asymptotic properties of the algorithm as the number of neurons grows. In particular, neural networks with a large number of parameters admit a mean-field description, which has recently served as a theoretical explanation for its favorable training properties. In this regime, gradient descent obeys a deterministic partial differential equation (PDE) that converges to a globally optimal solution for networks with a single hidden layer under appropriate assumptions.

In this talk, we will review recent progress on this problem, and will describe a non-local mass transport dynamics that leads to a modified PDE with the same minimizer. We implement this non-local dynamics as a stochastic neuronal birth-death process and we prove that it accelerates the rate of convergence in the mean-field limit. We subsequently realize this PDE with two classes of numerical schemes that converge to the mean-field equation, each of which can easily be implemented for neural networks with finite numbers of parameters. We will illustrate our algorithms with empirical examples to provide intuition for the mechanism through which convergence is accelerated, and discuss current open problems in this research direction.

Joint work with G. Rotskoff (NYU), S. Jelassi (Princeton) and E. Vanden-Eijnden (NYU).