

Global convergence of ResNets: From finite to infinite width using linear parameterization

Raphaël Barboni
École Normale Supérieure - PSL
raphael.barboni@ens.fr

Overparameterization is a key factor in the absence of convexity to explain global convergence of gradient descent (GD) for neural networks. Beside the well studied lazy regime, infinite width (mean field) analysis has been developed for shallow networks, using on convex optimization technics. To bridge the gap between the lazy and mean field regimes, we study Residual Networks (ResNets) in which the residual block has linear parameterization while still being nonlinear. Such ResNets admit both infinite depth and width limits, encoding residual blocks in a Reproducing Kernel Hilbert Space (RKHS). In this limit, we prove a local Polyak-Lojasiewicz inequality. Thus, every critical point is a global minimizer and a local convergence result of GD holds, retrieving the lazy regime. In contrast with other mean-field studies, it applies to both parametric and non-parametric cases under an expressivity condition on the residuals. Our analysis leads to a practical and quantified recipe: starting from a universal RKHS, Random Fourier Features are applied to obtain a finite dimensional parameterization satisfying with high-probability our expressivity condition.

Joint work with: Gabriel Peyré, François-Xavier Vialard.