From multiplicative matrix-valued diffusion to isometry of residual networks in deep learning

Maciej Nowak
Jagiellonian University, Cracow, Poland

We demonstrate that in residual neural networks (ResNets) dynamical isometry is achievable irrespective of the activation function used. We do that by deriving, with the help of Free Probability and Random Matrix Theories, a universal formula for the spectral density of the input-output Jacobian at initialization, in the large network width and depth limit. The resulting singular value spectrum depends on a single parameter, which we calculate for a variety of popular activation functions, by analyzing the signal propagation in the artificial neural network. We corroborate our results with numerical simulations of both random matrices and ResNets applied to the CIFAR-10 classification problem. Moreover, we study consequences of this universal behavior for the initial and late phases of the learning processes. We conclude by drawing attention to the simple fact, that initialization acts as a confounding factor between the choice of activation function and the rate of learning. We propose that in ResNets this can be resolved based on our results by ensuring the same level of dynamical isometry at initialization. The presentation is based on work [1].

References

[1] Dynamical Isometry is Achieved in Residual Networks in a Universal Way for any Activation Function, Wojciech Tarnowski, Piotr Warchoł, Stanisław Jastrzębski, Jacek Tabor, Maciej Nowak Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics, PMLR 89:2221-2230, 2019.