

# GRADIENT DESCENT AND STOCHASTIC GRADIENT DESCENT CONVERGENCE FOR LEARNING LINEAR NEURAL NETWORKS

**Gabin Maxime Nguegnang**  
RWTH Aachen University, Germany  
nguegnang@mathc.rwth-aachen.de

We study the convergence properties of gradient descent and stochastic gradient descent for learning deep linear neural networks. First of all, we extend a previous analysis for the related gradient flow. We show that under suitable conditions on the step sizes gradient descent converges to a critical point of the square loss function. Moreover, we prove that for almost all initialization gradient descent converges to a global minimum in the case of two layers. In the case of three or more layers we show that gradient descent converges to a global minimum on the manifold matrices of some fixed rank, where the rank cannot be determined a priori. Furthermore, We use an analytical approach that combines stochastic gradient descent iterates and gradient flow trajectories base on stochastic approximation theory to analyze stochastic gradient descent dynamic. Then establish the almost sure boundedness of stochastic gradient descent iterates and its convergence guarantee for learning deep linear neural networks. Most studies on the analysis of stochastic gradient descent for nonconvex problem have entirely focused on convergence property which only indicate that the second moment of the loss function gradient tend to zero. Our work demonstrates the convergence of stochastic gradient descent to a critical point of the square loss almost surely for learning deep linear neural networks.

*Joint work with Bubacarr Bah (Medical Research Council Unit The Gambia at the London School of Hygiene & Tropical Medicine, Gambia), Holger Rauhut (RWTH Aachen University, Germany) and Ulrich Terstiege (RWTH Aachen University, Germany).*