

# LINE SEARCH METHODS FOR DEEP LEARNING. IT COULD WORK

**Leonardo Galli**

RWTH Aachen University, Germany

galli@mathc.rwth-aachen.de

Stochastic Gradient Descent (SGD) is the horsepower for the whole deep learning activity today. Even though its simplicity and low memory requirements seem crucial for dealing with these huge models, the success of SGD is deeply connected to the choice of the learning rate. In this paper we show that line search methods are a valid alternative to SGD for training convolutional deep learning models and transformers. Following the classical optimization doctrine, we here combine a fast initial step size (Polyak) with a nonmonotone line search. We show that to achieve the best performances, the initial step size sometimes needs a line search to control its growth, especially while still in the global phase. This behavior agrees with common optimization knowledge and with the theoretical expectations regarding line search methods. Moreover, to deal with the increased amount of backtracking steps, we here develop a new technique that is able to reduce them to 0 (in average), while not altering the behavior of the original step size. To conclude, we prove the first rates of convergence for nonmonotone line search methods in the stochastic setting under interpolation.

*Joint work with Holger Rauhut (RWTH Aachen University) and Mark Schmidt (University of British Columbia).*