



UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

Dipartimento di Scienze Fisiche,
Informatiche e Matematiche

Martedì 1 settembre 2026 · ore 15:30

Dipartimento FIM · Via Campi 213/b, Modena – 41125 · Laboratorio Zironi

Dr. Leonardo Galli
LMU Munich

Flatland: The Adventures of Gradient Descent with Large Step Sizes

Abstract

The training of neural networks often entails objective functions that are not globally L -smooth. For these functions, it is both theoretically and practically difficult to answer the question: what is the largest possible step size that ensures the convergence of gradient descent (GD)? We address this longstanding open question in deep learning by providing a unifying definition of "large" step sizes that requires only local Lipschitz (or even Hölder) continuity of the gradient. We design first-order adaptive methods that provably yield large step sizes and show that they operate at the edge of stability (EoS) right from the start of the training. In particular, the loss decreases nonmonotonically and the product between the step size and sharpness, i.e., the largest eigenvalue of the Hessian, stays above the EoS threshold of 2 throughout training. Using our method, we are able to decrease the sharpness down to its global minimum. Contrary to expectation, we find that encountering globally-flat regions too early in training may slow down convergence and jeopardize the generalization ability of the network. Exploiting a self-stabilization argument, we allow GD to enter slightly sharper valleys and turn unsuccessful training runs into successful ones.