

# TWO SIDES OF ONE COIN: THE LIMITS OF UNTUNED SGD AND THE POWER OF ADAPTIVE METHODS

**Niao He**

ETH Zurich , Switzerland  
niao.he@inf.ethz.ch

A central optimization challenge in machine learning is parameter-tuning. Adaptive gradient methods, such as AdaGrad and Adam, are ubiquitously used for training machine learning models in practice, owing to their ability to adjust the stepsizes without granular knowledge of the objective functions. Despite the empirical successes, their theoretical benefits over vanilla SGD remain elusive. In this talk, we will examine the convergence guarantees of a wide range of parameter-agnostic algorithms including untuned SGD, normalized SGD, AdaGrad, and others in the nonconvex setting assuming only smoothness and bounded variance. Our results will provide some hints on the provable advantage of adaptive methods.