Adaptive Step Sizes for Preconditioned Stochastic Gradient Descent

Monday, August 12, 2024 4:30 PM (30 minutes)

The choice of the step size (or learning rate) in stochastic optimization algorithms, such as stochastic gradient descent, plays a central role in the training of machine learning models. Both theoretical investigations and empirical analyses emphasize that an optimal step size not only requires taking into account the nonlinearity of the underlying problem, but also relies on accounting for the local variance within the search directions. In this presentation, we introduce a novel method capable of estimating these fundamental quantities and subsequently using these estimates to derive an adaptive step size for stochastic gradient descent. Our proposed approach leads to a nearly hyperparameter-free variant of stochastic gradient descent. We provide theoretical convergence analyses in the special case of stochastic quadratic, strongly convex problems. In addition, we perform numerical experiments focusing on classical image classification tasks. Remarkably, our algorithm exhibits truly problem-adaptive behavior when applied to these problems that exceed theoretical boundaries. Moreover, our framework facilitates the potential incorporation of a preconditioner, thereby enabling the implementation of adaptive step sizes for stochastic second-order optimization methods.

Authors: Prof. SCHIELA, Anton (Universität Bayreuth); KÖHNE, Frederik (Universität Bayreuth); KREIS, Leonie (Heidelberg University); HERZOG, Roland (Heidelberg University)

Presenter: KÖHNE, Frederik (Universität Bayreuth)

Session Classification: MS 01: Optimal Control and Machine Learning

Track Classification: Minisymposia: MS 01: Optimal Control and Machine Learning