Variance-Reduction for SGD with Polyak Stepsizes and Line-Search

Xiaowen Jiang

CISPA Helmholtz Center for Information Security, Germany
xiaowenjiang98@gmail.com

The recently proposed stochastic Polyak stepsize (SPS) and stochastic line-search (SLS) for SGD have shown remarkable effectiveness when training over-parameterized models. However, in non-interpolation settings, both algorithms only guarantee convergence to a neighborhood of a solution which may result in worse output than the initial guess. While artificially decreasing the adaptive stepsize has been proposed to address this issue (Orvieto et al. [2022]), this approach results in slower convergence rates for convex and over-parameterized models.

In this work, we make two contributions: Firstly, we propose two new variants of SPS and SLS, called AdaSPS and AdaSLS, which guarantee convergence in non-interpolation settings and maintain sub-linear and linear convergence rates for convex and strongly convex functions when training over-parameterized models. These algorithms require no knowledge of problem-dependent parameters and do not assume bounded iterates under strong convexity assumption. Secondly, we equip AdaSPS and AdaSLS with a novel variance reduction technique and obtain algorithms that require $\widetilde{\mathcal{O}}(n + 1/\epsilon)$ gradient evaluations to achieve an $\mathcal{O}(\epsilon)$-suboptimality for convex functions, which improves upon the slower $\mathcal{O}(1/\epsilon^2)$ rates of AdaSPS and AdaSLS without variance reduction in the non-interpolation regimes. Additionally, our result matches the fast rates of AdaSVRG and SARAH.

Joint work with Sebastian U. Stich (CISPA Helmholtz Center for Information Security).