AHMET ALACAOGLU, UBC

Towards Weaker Variance Assumptions for Stochastic Optimization: A Blast From the Past

In this talk, we focus on a classical assumption for analyzing stochastic gradient algorithms where the squared norm of the stochastic subgradient (or the variance for smooth problems) is allowed to grow as fast as the squared norm of the optimization variable. We contextualize this assumption in view of its inception in the 1960s, its seemingly independent appearance in the recent literature, its relationship to weakest-known variance assumptions for analyzing stochastic gradient algorithms, and its relevance even in deterministic problems for non-Lipschitz nonsmooth convex optimization. We build on and extend a connection recently made between this assumption and the Halpern iteration in view of nonasymptotic convergence rates for stochastic optimization. For convex nonsmooth, and potentially stochastic, optimization we provide horizon-free algorithms with last-iterate rates. For problems beyond simple constrained optimization, such as convex problems with functional constraints, we obtain rates for optimality measures that do not require boundedness of the feasible set. (Joint work with Yura Malitsky and Stephen J. Wright)