Semi Stochastic Gradient Descent

Jakub Konečný
University of Edinburgh
kubo.konecny@gmail.com

In this work we study the problem of minimizing the average of a large number (n) of smooth convex loss functions. We propose a new method, S2GD (Semi-Stochastic Gradient Descent), which runs for one or several epochs in each of which a single full gradient and a random number of stochastic gradients is computed, following a geometric law. The total work needed for the method to output an ε-accurate solution in expectation, measured in the number of passes over data, or equivalently, in units equivalent to the computation of a single gradient of the loss, is $O(\log(1/\varepsilon))$. This is achieved by running the method for $O(\log(1/\varepsilon))$ epochs, with a single gradient evaluation and $O(\kappa)$ stochastic gradient evaluations in each, where κ is condition number of the problem. The SVRG method of Johnson and Zhang (SVRG) arises as a special case. To illustrate our theoretical results, S2GD only needs the workload equivalent to about 2.1 full gradient evaluations to find an 10^{-6}-accurate solution for a problem with $n = 10^9$ and $\kappa = 10^3$. Furthermore, we present a minibatching scheme, which admits simple possibility of parallelism and even improves the complexity bound under certain conditions.

Joint work with Peter Richtárik.