Accelerating Stochastic Gradient Descent

Kanan Gupta
Texas A&M University, United States of America
kanan@tamu.edu

Momentum-based gradient descent methods use information gained along the trajectory, in addition to the local information from the gradient, in order to achieve an accelerated rate of convergence. These methods have been well-studied for convex optimization. Computing the gradient is often too expensive and it is approximated using stochastic gradient estimates in practice. However, there's a lack of theoretical analyses of accelerated methods in the setting of stochastic gradient descent, even for the simple case of convex functions. We address this gap with a novel descent algorithm which provably achieves the optimal convergence rate for convex optimization. While the objective functions in deep learning training are non-convex, they share many properties with convex functions. Empirical results show that our algorithm outperforms the existing variants of stochastic gradient descent with momentum for training of neural networks.

Joint work with Jonathan Siegel (Texas A&M University) and Stephan Wojtowytsch (Texas A&M University).