Asymptotics for Constant Step Size Stochastic Gradient Descent

Wei Biao Wu

Department of Statistics, University of Chicago

I will discuss a novel approach to understanding the behavior of Stochastic Gradient Descent (SGD) with constant step size by interpreting its evolution as a Markov chain. Unlike previous studies that rely on the Wasserstein distance, our approach leverages the functional dependence measure and explore the Geometric-Moment Contraction (GMC) property to capture the general asymptotic behavior of SGD in a more refined way. In particular, our approach allow SGD iterates to be non-stationary but asymptotically stationary over time, providing quenched versions of the central limit theorem and invariance principle valid for averaged SGD with any given starting point. These asymptotic results allow for the initialization of SGD with multiple distinct step sizes, which is a widespread practice in the discipline. We subsequently show a Richardson-Romberg extrapolation with an improved bias representation to bring the estimates closer to the global optimum. We establish the existence of a stationary solution for the derivative SGD process under mild conditions, enhancing our understanding of the entire SGD procedure across varied step sizes. Lastly, we propose an efficient online method for estimating the long-run variance of SGD solutions. This aligns with the recursive nature of SGD, thereby facilitating fast and efficient computations. The work is joint with Jiaqi Li, Zhipeng Lou and Stefan Richter.