

---

**MARK SCHMIDT**, University of British Columbia  
*Faster Algorithms for Deep Learning?*

The last 10 years have seen a revolution in stochastic gradient methods, with variance-reduced methods like SAG/SVRG provably achieving faster convergence rates than all previous methods. These methods give dramatic speedups in a variety of applications, but have had virtually no impact to the practice of training deep models. We hypothesize that this is due to the over-parameterized nature of modern deep learning models, where the models are so powerful that they could fit every training example with zero error (at least theoretically). Such over-parameterization nullifies the benefits of variance-reduced methods, because in some sense it leads to "easier" optimization problems. In this work, we present algorithms specifically designed for over-parameterized models. This leads to methods that provably achieve Nesterov acceleration, methods that automatically tune the step-size as they learn, and methods that achieve superlinear convergence with second-order information.