

1139-90-362

**Sanghamitra Dutta, Gauri Joshi, Soumyadip Ghosh\*** (ghoshs@us.ibm.com), **Parijat Dube** and **Priya Nagpurkar**. *Slow and Stale Gradients Can Win the Race: Error-Runtime Trade-offs in Distributed SGD*.

Distributed Stochastic Gradient Descent (SGD) when run in a synchronous manner, suffers from delays in waiting for the slowest learners (stragglers). Asynchronous methods can alleviate stragglers, but cause gradient staleness that can adversely affect convergence. In this work we present the first theoretical characterization of the speed-up offered by asynchronous methods by analyzing the trade-off between the error in the trained model and the actual training runtime (wallclock time). The novelty in our work is that our runtime analysis considers random straggler delays, which helps us design and compare distributed SGD algorithms that strike a balance between stragglers and staleness. We also present a new convergence analysis of asynchronous SGD variants without bounded or exponential delay assumptions, and a novel learning rate schedule to compensate for gradient staleness. (Received February 16, 2018)