

Learning under Delayed Feedback: Implicitly Adapting to Gradient Delays

Rotem Aviv Ido Hakimi Assaf Shuster Kfir Y. Levy

Abstract

We consider stochastic convex optimization problems, where several machines act asynchronously in parallel while sharing a common memory. We propose a robust training method for the constrained setting and derive non asymptotic convergence guarantees that do not depend on prior knowledge of update delays, objective smoothness, and gradient variance. Conversely, existing methods for this setting crucially rely on this prior knowledge, which render them unsuitable for essentially all shared-resources computational environments, such as clouds and data centers. Concretely, existing approaches are unable to accommodate changes in the delays which result from dynamic allocation of the machines, while our method implicitly adapts to such changes.