An Even More Optimal Stochastic Optimization Algorithm: Minibatching and Interpolation Learning

Anonymous Author(s)
Affiliation
Address
email

Abstract

We present and analyze an algorithm for optimizing smooth and convex or strongly convex objectives using minibatch stochastic gradient estimates. The algorithm is optimal with respect to its dependence on both the minibatch size and minimum expected loss simultaneously. This improves over the optimal method of Lan [2], which is insensitive to the minimum expected loss; over the optimistic acceleration of Cotter et al. [1], which has suboptimal dependence on the minibatch size; and over the algorithm of Liu and Belkin [3], which is limited to least squares problems and is also similarly suboptimal. Applied to interpolation learning, the improvement over Cotter et al. and Liu and Belkin translates to a linear, rather than square-root, parallelization speedup.

1 Introduction

References

[1] Andrew Cotter, Ohad Shamir, Nati Srebro, and Karthik Sridharan. Better mini-batch algorithms via accelerated gradient methods. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 1647–1655. Curran Associates, Inc., 2011. URL `http://papers.nips.cc/paper/4432-better-mini-batch-algorithms-via-accelerated-gradient-methods.pdf`.

[2] Guanghui Lan. An optimal method for stochastic composite optimization. *Mathematical Programming*, 133(1-2):365–397, 2012. URL `https://pdfs.semanticscholar.org/1621/f05894ad5fd6a8fcb8827a8c7aca36c81775.pdf`.

[3] Chaoyue Liu and Mikhail Belkin. Mass: an accelerated stochastic method for over-parametrized learning. *arXiv preprint arXiv:1810.13395*, 2018.

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.