Communication Efficient and Differentially Private Optimization

Shuli Jiang

Thesis Proposal

November 11, 2024

The Robotics Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA

Thesis Committee:
Gauri Joshi, *chair*
Steven Wu
Zachary Manchester
Swanand Kadhe, *IBM Research*

*Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Robotics.*

Abstract

In recent years, the integration of communication efficiency and differential privacy in distributed optimization has gained significant attention, motivated by large-scale applications such as Federated Learning (FL), where both data privacy and efficient communication are critical. This thesis explores the development of novel techniques to address these challenges, with a focus on distributed mean estimation, differentially private prediction, and private optimization for empirical risk minimization.

The first part of this work addresses communication-efficient distributed vector mean estimation, an essential subroutine in distributed optimization and FL. We propose the Rand-Proj-Spatial family estimator which utilizes cross-client correlation to reduce the estimation error under fixed communication cost, by projecting client vectors into a random subspace using a Subsampled Randomized Hadamard Transform (SRHT). This approach captures cross-client correlation more effectively, demonstrating substantial performance gains over conventional sparsification techniques in various distributed optimization tasks.

The second part of this work focuses on maximizing the privacy-utility trade-offs in differentially private prediction through majority ensembling. We introduce the Data-dependent Randomized Response Majority (DaRRM) framework, which generalizes all private majority ensembling algorithms through a data-dependent noise function. Based on DaRRM, we propose a computationally tractable optimization procedure for maximizing utility under a fixed privacy loss. Empirical results demonstrate DaRRM's effectiveness in private label ensembling for image classification, showing significant utility improvements over existing baselines.

The third part of this work investigates differentially private optimization in solving empirical risk minimization using shuffled gradient methods. Unlike conventional private optimizers such as DP-SGD, which benefits from privacy amplification by subsampling, shuffled gradient methods face unique challenges in privacy and convergence. We develop a theoretical framework for analyzing Incremental Gradient (IG) methods, the most basic form of shuffled gradient methods, that enables noise injection for privacy and the use of surrogate objectives, introducing a new dissimilarity metric to measure the difference between true and surrogate objectives. Leveraging privacy amplification by iteration, we establish the first empirical excess risk bound for differentially private IG (DP-IG), and show how interleaving public data in training can further improve privacy-convergence trade-offs in DP-IG.

Finally, we introduce two proposed works along the line of differentially private optimization. First, we aim to extend our theoretical framework to analyze Shuffle Once (SO) and Random Reshuffling (RR), two practical shuffled gradient methods beyond Incremental Gradient (IG) methods. This will enable us to understand their private counterparts, DP-SO and DP-RR, where privacy analysis is more complex due to a lack of understanding on privacy amplification through shuffling. Second, we plan to extend our framework from a local to a distributed or decentralized setting to analyze convergence rates of distributed shuffled gradient methods in both private and non-private contexts, while also investigating the impact of data heterogeneity among clients on convergence in this setting.