CSIAM Trans. Appl. Math. Vol. **6**, No. 2, pp. 207-249
doi: 10.4208/csiam-am.SO-2024-0023 June 2025

Review of Mathematical Optimization in Federated Learning

Shusen Yang[1,2], Fangyuan Zhao[2,3], Zihao Zhou[1,2], Liang Shi[2,3],
Xuebin Ren[2,3,*] and Zongben Xu[1,2]

[1] *School of Mathematics and Statistics,Xi'an Jiaotong University, Xi'an 710049, China.*
[2] *National Engineering Laboratory for Big Data Analytics, Xi'an Jiaotong University, Xi'an 710049, China.*
[3] *Faculty of Electronic and Information, Xi'an Jiaotong University, Xi'an 710049, China.*

Received 7 June 2024; Accepted 1 December 2024

Abstract. Federated learning (FL) has been becoming a popular interdisciplinary research area in both applied mathematics and information sciences. Mathematically, FL aims to collaboratively optimize aggregate objective functions over distributed datasets while satisfying a variety of privacy and system constraints. Different from conventional distributed optimization methods, FL needs to address several specific issues (e.g. non-i.i.d. data and differential private noises), which pose a set of new challenges in the problem formulation, algorithm design, and convergence analysis. In this paper, we will systematically review existing FL optimization research including their assumptions, formulations, methods, and theoretical results. Potential future directions are also discussed.

AMS subject classifications: 90C26, 90C31, 68W15, 68T05

Key words: Federated learning, distributed optimization, convergence analysis, error bounds.

1 Introduction

With the increasingly stringent privacy regulations [145, 180], data isolation has been becoming the key bottleneck of data sciences and artificial intelligence. To address this issue, federated learning emerges as a popular privacy-preserving distributed machine learning (ML) paradigm, which enables multiple data owners to jointly train ML mod-

*Corresponding author. *Email addresses:* `shusenyang@mail.xjtu.edu.cn` (S. Yang), `zfy1454236335@stu.xjtu.edu.cn` (F. Zhao), `wszzh15139520600@stu.xjtu.edu.cn` (Z. Zhou), `sl1624@stu.xjtu.edu.cn` (L. Shi), `xuebinren@mail.xjtu.edu.cn` (X. Ren), `zbxu@mail.xjtu.edu.cn` (Z. Xu)

els without sharing the raw data [71, 91, 92]. It has gained extensive interests from both academia and industry, and demonstrated great success across multiple domains, including medicine [48], finance [206], and industry [228].

Mathematically, FL training tasks are essentially distributed optimization problems, which aim to minimize aggregate global objectives (e.g. the mean empirical loss), across a set of distributed data owners by exchanging model parameters trained on their local datasets [178, 231, 232]. Despite being similar in essence, FL optimization has many distinct characteristics from traditional distributed optimization. Their main difference lies in the communication environment. Specifically, traditional distributed optimization, mainly in the form of distributed ML [18, 107, 139] (some used in distributed resource allocation [9, 13, 72]), is often used for high-throughput ML speedup in data centers. Here, multiple homogeneous computing nodes with uniformly distributed data partitions are connected by reliable networks with gigabytes of bandwidth. However, FL is often applied to achieve collaborative and privacy-preserving ML in wide-area networks, where geographically distributed clients with naturally generated data collaborate to train ML models over bandwidth-constrained communication channels.

Due to the drastic differences, FL needs to address a variety of specific and complicated issues in optimization. For example, naturally generated data at the FL clients are commonly heterogeneous, i.e. non-balanced and non-i.i.d. distributed, which leads to biased model aggregation. Considering limited communication bandwidth, FL often performs multiple local updates before the global aggregation, further amplifying the model bias caused by data heterogeneity. Furthermore, model dissemination and aggregation give rise to concerns of private information leakage and falsification. The decentralized communication architecture also leads to partially local approximations of global objective functions. These issues pose new challenges in optimization in FL, including problem formulation, algorithm design, and convergence analysis. In particular, typical challenges are summarized as follows:

- **Biased local objectives from non-i.i.d. datasets.** The potentially unknown and non-i.i.d. data among distributed nodes[†] could result in biased local objectives [91, 239]. Thus, the global objective function in FL optimization cannot be decomposed into trainable local objectives without bias. This means that the gradients of local models may largely deviate from the steepest descent direction of the global objective, leading to significant degradation of convergence speed and model accuracy [97].

- **Perturbed gradients with DP noises.** The differential privacy (DP) mechanisms [51] are commonly adopted in the FL optimization process to protect exchanged parameters or gradients, by introducing statistically unbiased noises (e.g. Gaussian [1] or Laplacian [198] noises). Despite the statistical unbiasedness, random

[†]The terms "node", "client", "participant" and "data owner" are used interchangeably within this survey.