Sign SGD with Heavy-Tailed Noise and Differential Privacy

Alexey Kravatskiy
kravtskii.aiu@phystech.edu

Anton Plusnin
plusnin.aa@phystech.edu

Savelii Chezhegov
chezhegov.sa@phystech.edu

March 5, 2025

Abstract

Federated learning faces the dual challenge of ensuring fast communication and maintaining data privacy. This paper addresses these challenges by proposing a novel modification of the Sign-SGD algorithm, designed to be both communication-efficient and differentially private. Our approach is particularly suited for real-world data characterized by noise distributions with infinite variance and bounded k-th moments, where k ranges from 1 to 2. We demonstrate that our modified Sign-SGD algorithm achieves differential privacy and converges with high probability, even when the noise is heavy-tailed, asymmetric, or non-unimodal. Additionally, we show that the algorithm's speed is comparable to existing methods. The effectiveness of our approach is validated through improved training quality and speed in Large Language Models.

Keywords: Sign SGD, differential privacy, high-probability convergence, federated learning, heavy-tailed noise.

1 Introduction

In the recent work [2], the authors have shown that Sign-SGD with heavy-tailed noise achieves optimal convergence rates under certain assumptions. Additionally, [1] demonstrated that Sign-SGD can be made differentially private while maintaining good convergence properties.

2 Theory

Present your theoretical framework, definitions, lemmas, and proofs.

3   Experiments

Describe your experimental setup, methodology, and results.

4   Conclusion

Summarize your findings and discuss future work.

Acknowledgments

Optional acknowledgments section.

A   Additional Proofs and Results

Include detailed proofs and supplementary materials here.

References

[1] Richeng Jin et al. "Stochastic-Sign SGD for Federated Learning with Theoretical Guarantees". In: *Part of this work is published in IEEE Transactions on Neural Networks and Learning Systems, 2024* 36.2 (Feb. 25, 2020), pp. 3834–3846. ISSN: 2162-2388. DOI: 10.1109/tnnls.2023.3345367. arXiv: 2002.10940 [cs.LG].

[2] Nikita Kornilov et al. *Sign Operator for Coping with Heavy-Tailed Noise: High Probability Convergence Bounds with Extensions to Distributed Optimization and Comparison Oracle.* 2025. DOI: 10.48550/ARXIV.2502.07923.