Sign operator for (L_0, L_1)-smooth optimization

Mark Ikonnikov
ikonnikov.mi@phystech.edu

Nikita Kornilov
kornilov.nm@phystech.edu

March 5, 2025

Abstract

In Machine Learning, the non-smoothness of optimization problems, the high cost of communicating gradients between workers, and severely corrupted data during training necessitate generalized optimization approaches. This paper explores the efficacy of sign-based methods [1], which address slow transmission by communicating only the sign of each minibatch stochastic gradient. We investigate these methods within (L_0, L_1)-smooth problems [2], which encompass a wider range of problems than the L-smoothness assumption. Furthermore, under the assumptions above, we investigate techniques to handle heavy-tailed noise [4], defined as noise with bounded κ-th moment $\kappa \in (1, 2]$. This includes the use of SignSGD with Majority Voting in the case of symmetric noise. We then attempt to extend the findings to convex cases using error feedback [3].

Keywords: Sign-based methods, (L_0, L_1)-smoothness, high-probability convergence, heavy-tailed noise.

Highlights below to be fixed later (these are our hopes for the paper)

Highlights:

1. Proves convergence of sign-based methods for (L_0, L_1)-smooth optimization
2. Handles heavy-tailed noise with high-probability convergence guarantees
3. Extends sign-based optimization to convex functions using error feedback

1 Introduction

TODO

References

[1] Jeremy Bernstein et al. "signSGD: Compressed Optimisation for Non-Convex Problems". In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. PMLR, Oct. 2018, pp. 560–569. URL: https://proceedings.mlr.press/v80/bernstein18a.html.

[2] Eduard Gorbunov et al. *Methods for Convex (L_0, L_1)-Smooth Optimization: Clipping, Acceleration, and Adaptivity*. 2024. arXiv: 2409.14989 [math.OC]. URL: https://arxiv.org/abs/2409.14989.

[3] Sai Praneeth Karimireddy et al. "Error Feedback Fixes SignSGD and other Gradient Compression Schemes". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, Sept. 2019, pp. 3252–3261. URL: https://proceedings.mlr.press/v97/karimireddy19a.html.

[4] Kornilov Nikita 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.