# Sign Operator for $(L_0, L_1)$-Smooth Optimization with Heavy-Tailed Noise


Ikonnikov Mark

Moscow Institute of physics and Technology


*Course:* My first scientific paper (Strijov's practice)/Group 206

*Expert:* Alexander Beznosikov

*Consultant:* Nikita Kornilov

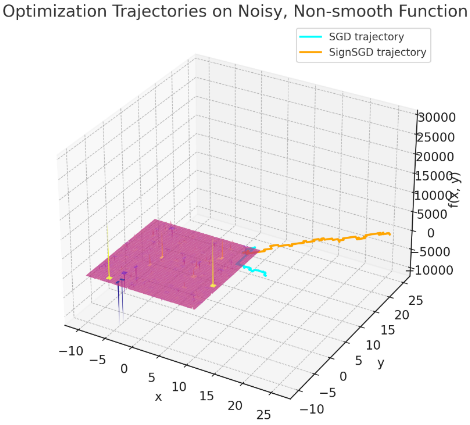
April 17, 2025

# Goal of Research

## Objectives

- Define $(L_0, L_1)$-smoothness.
- Develop sign-based methods (Sign-SGD, minibatch-SignSGD, momentum-SignSGD) for heavy-tailed (HT) noise.
- Establish theoretical convergence bounds under $(L_0, L_1)$-smoothness and HT noise.
- Validate results through computational experiments.

# Idea




Convergence rates improve significantly with Sign-methods.

$\|\nabla^2 f(x)\|_2 \leq L_0 + L_1\|\nabla f(x)\|$

$\mathbb{E}_\xi[|\nabla f(x,\xi)_i - \nabla f(x)_i|^\kappa] \leq \sigma_i^\kappa,\ \kappa \in (1,2]$

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

# Literature

| Title | Year | Authors | Paper |
|---|---|---|---|
| Sign Operator for Coping with Heavy-Tailed Noise | 2025 | Kornilov et al. | arXiv |
| signSGD: Compressed Optimisation for Non-Convex Problems | 2018 | J. Bernstein et al. | PMLR |
| Methods for Convex (L0,L1)-Smooth Optimization | 2024 | Gorbunov et al. | arXiv |
| Robustness to Unbounded Smoothness of Generalized SignSGD | 2022 | M. Crawshaw et al. | NeurIPS |

# Hypothesis and Model

## Hypothesis

Sign-based optimization methods outperform traditional gradient-based methods in $(L_0, L_1)$-smooth problems with heavy-tailed noise, achieving faster convergence and robustness.

## Model

$f : \mathbb{R}^d \to \mathbb{R}$:

$$\min_{x \in \mathbb{R}^d} f(x) := \mathbb{E}_{\xi \sim \mathcal{S}}[f(x, \xi)],$$

Consider a function $f : \mathbb{R}^d \to \mathbb{R}$ that is $(L_0, L_1)$-smooth:

$$\|\nabla f(x) - \nabla f(y)\| \leq (L_0 + L_1 \|\nabla f(u)\|) \|x - y\|,$$

with gradient estimates $\nabla f(x, \xi)$ under HT noise:

- $\mathbb{E}_\xi[\nabla f(x, \xi)] = \nabla f(x)$,
- $\mathbb{E}_\xi[|\nabla f(x, \xi)_i - \nabla f(x)_i|^\kappa] \leq \sigma_i^\kappa$, $\kappa \in (1, 2]$.

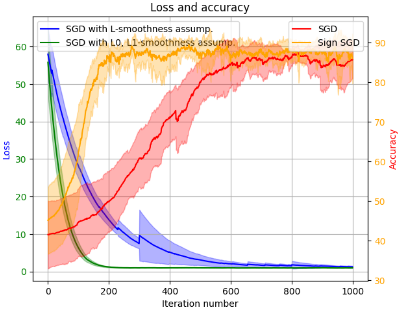
# Solution: Theoretical Part

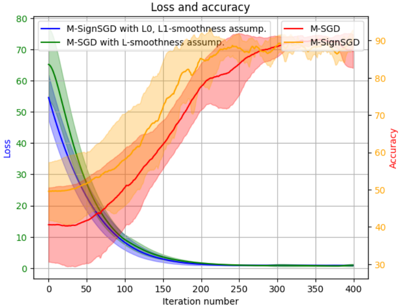
**Theorem (HP complexity for minibatch-L0L1-SignSGD)**

*Consider lower-bounded $(L0, L1)$-smooth function $f$ and HT gradient estimates. Then Alg. minibatch-SignSGD requires the sample complexity $N$ to achieve $\frac{1}{T}\sum_{k=1}^{T}\|\nabla f(x^k)\|_1 \le \varepsilon$ with probability at least $1-\delta$ for:*

**Optimal tuning.** *In case $\varepsilon \ge \frac{8L_0}{L_1\sqrt{d}}$, we use stepsize $\gamma = \frac{1}{48L_1 d \log\frac{1}{\delta}\sqrt{d}} \Rightarrow 80L_0 d\gamma\log(1\delta) \le \varepsilon/2$ and batchsize $B_k \equiv \max\left\{1, \left(\frac{16\|\vec{\sigma}\|_1}{\varepsilon}\right)^{\frac{\kappa}{\kappa-1}}\right\}$. $T = O\left(\frac{\Delta_1 L_1 \log\frac{1}{\delta} d^{\frac{3}{2}}}{\varepsilon}\right)$. The total number of oracle calls is:*

$$
\begin{aligned}
\varepsilon \;\ge\; \frac{8L_0}{L_1\sqrt{d}} \quad &\Rightarrow \quad N = O\left(\frac{\Delta_1 L_1 \log(1\delta) d^{\frac{3}{2}}}{\varepsilon}\left[1+\left(\frac{\|\vec{\sigma}\|_1}{\varepsilon}\right)^{\frac{\kappa}{\kappa-1}}\right]\right),\\
\varepsilon \;<\; \frac{8L_0}{L_1\sqrt{d}} \quad &\Rightarrow \quad N = O\left(\frac{\Delta_1 L_0 \log(1\delta) d}{\varepsilon^2}\left[1+\left(\frac{\|\vec{\sigma}\|_1}{\varepsilon}\right)^{\frac{\kappa}{\kappa-1}}\right]\right).
\end{aligned}
$$




Convergence and accuracy rates improve significantly with Sign-methods.

# Error Analysis

## Error comparison

| Method | Mean Loss | Mean Acc. | Loss Var. | Acc. Var. |
|---|---|---|---|---|
| M-SignSGD | 3.63 | 82.86 | 73.56 | 135.77 |
| M-SGD | 7.72 | 73.46 | 209.46 | 341.58 |
| SignSGD | 6.71 | 79.12 | 155.10 | 140.47 |
| SGD | 16.44 | 62.96 | 234.20 | 70.55 |

Table: comparison of convergence of several methods under the assumptions

# Results and Conclusions

## Results

- Sign-based methods outperform SGD in convergence under $(L_0, L_1)$-smoothness and HT noise.
- minibatch-SignSGD reduces sample complexity for $\kappa < 2$.
- Momentum-SignSGD and minibatch-SignSGD convergence are bounded and proved.

## Conclusions

- $(L_0, L_1)$-smoothness enables better rates under $(L_0, L_1)$ and HT-noise assumptions.
- Sign-based methods are noise-robust and communication-efficient.