Differentially private modification of sign-SGD

Alexey Kravatskiy

Moscow Institute of Physics and Technology

Course: My first scientific paper (Strijov's practice)/Group 205

Expert: A. N. Beznosikov

Consultant: S. A. Chezhegov

2025

Distributed, Private, and Noise-resistant

Goal

A communication-efficient and private algorithm for distributed optimization converging under heavy-tailed noise (noise with unbounded variance).

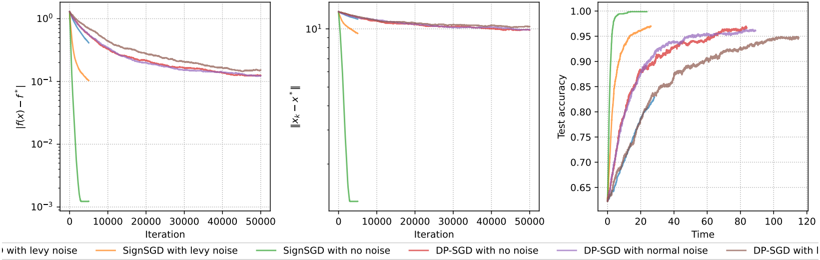
Problem

The only proposed private sign-based algorithm dp-signSGD either is not private or does not converge.

Solution

A new privacy accountant for dp-signSGD that affords lower noise to ensure privacy.

Differential privacy of DP-signSGD



DP-signSGD with an appropriate σ is $(\epsilon, \mathit{delta})$-private and converges under heavy-tailed noise.

Literature

- Jin, Richeng et al. (Feb. 25, 2020). "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, pp. 3834–3846. ISSN: 2162-2388. DOI: 10.1109/tnnls.2023.3345367. arXiv: 2002.10940 [cs.LG].
- Mironov, Ilya (Aug. 2017). "Rényi Differential Privacy". In: *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*. IEEE, pp. 263–275. DOI: 10.1109/csf.2017.11. URL: http://dx.doi.org/10.1109/CSF.2017.11.
- Mironov, Ilya, Kunal Talwar, and Li Zhang (2019). *Rényi Differential Privacy of the Sampled Gaussian Mechanism*. arXiv: 1908.10530 [cs.LG]. URL: https://arxiv.org/abs/1908.10530.

Differential Privacy

Definition

Given a set of local datasets $\mathcal{D}$ provided with a notion of neighboring local datasets $\mathcal{N}_{\mathcal{D}} \subset \mathcal{D} \times \mathcal{D}$ that differ in only one data point. For a query function $f : \mathcal{D} \to \mathcal{X}$, a mechanism $\mathcal{M} : \mathcal{X} \to \mathcal{O}$ to release the answer of the query is defined to be (ϵ, δ)-locally differentially private if for any measurable subset $\mathcal{S} \subseteq \mathcal{O}$ and two neighboring local datasets $(D_1, D_2) \in \mathcal{N}_{\mathcal{D}}$,

$$P(\mathcal{M}(f(D_1)) \in \mathcal{S}) \leq e^{\epsilon} P(\mathcal{M}(f(D_2)) \in \mathcal{S}) + \delta. \tag{1}$$

A key quantity in characterizing local differential privacy for many mechanisms is the sensitivity of the query f in a given norm l_r, which is defined as

$$\Delta_r = \max_{(D_1, D_2) \in \mathcal{N}_{\mathcal{D}}} ||f(D_1) - f(D_2)||_r. \tag{2}$$

Rényi divergence (Mironov 2017)

Definition (Rényi divergence)

Let P and Q be two distributions on $\mathcal{X}$ defined over the same probability space, and let p and q be their respective densities. The Rényi divergence of a finite order $\alpha \neq 1$ between P and Q is defined as

$$\mathrm{D}_\alpha(P \parallel Q) \triangleq \frac{1}{\alpha - 1} \ln \int_{\mathcal{X}} q(x) \left(\frac{p(x)}{q(x)} \right)^{\alpha} \mathrm{d}x.$$

Rényi divergence at orders $\alpha = 1, \infty$ are defined by continuity.

Rényi differential privacy (Mironov 2017)

Definition (Rényi differential privacy (RDP))

We say that a randomized mechanism $\mathcal{M} \colon \mathcal{S} \to \mathcal{R}$ satisfies (α, ε)-Rényi differential privacy (RDP) if for any two *adjacent* inputs $S, S' \in \mathcal{S}$ it holds that

$$\mathrm{D}_\alpha\big(\mathcal{M}(S) \,\|\, \mathcal{M}(S')\big) \leq \varepsilon.$$

Sample Gaussian Mechanism (Mironov, Talwar, and Zhang 2019)

Definition (Sampled Gaussian Mechanism (SGM))

Let f be a function mapping subsets of $\mathcal{S}$ to $\mathbb{R}^d$. We define the Sampled Gaussian mechanism (SGM) parameterized with the sampling rate $0 < q \leq 1$ and the noise $\sigma > 0$ as

$$\mathrm{SG}_{q,\sigma}(S) \triangleq f(\{x \colon x \in S \text{ is sampled with probability } q\}) + \mathcal{N}(0, \sigma^2 \mathbb{I}^d),$$

where each element of S is sampled independently at random with probability q without replacement, and $\mathcal{N}(0, \sigma^2 \mathbb{I}^d)$ is spherical d-dimensional Gaussian noise with per-coordinate variance σ^2.

Criterion of a private algorithm

Following the procedure from Mironov, Talwar, and Zhang 2019, we can get:

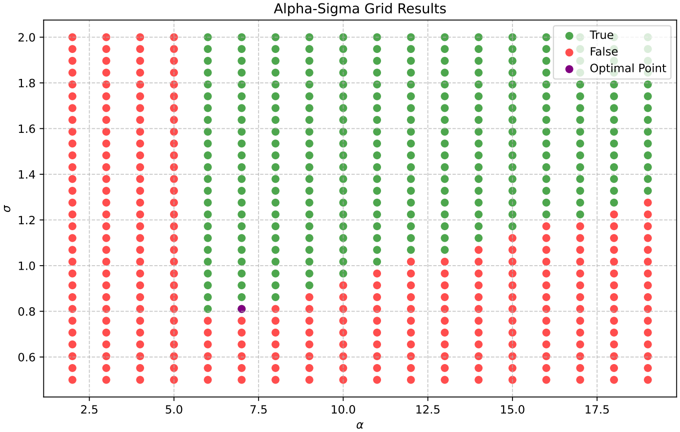
$$\varepsilon_R \leq \frac{1}{\alpha - 1} \log \left(\sum_{k=0}^{\alpha} \binom{\alpha}{k} (1-q)^{\alpha-k} q^k \exp\left(\frac{k^2 - k}{2\sigma^2} \right) \right)$$

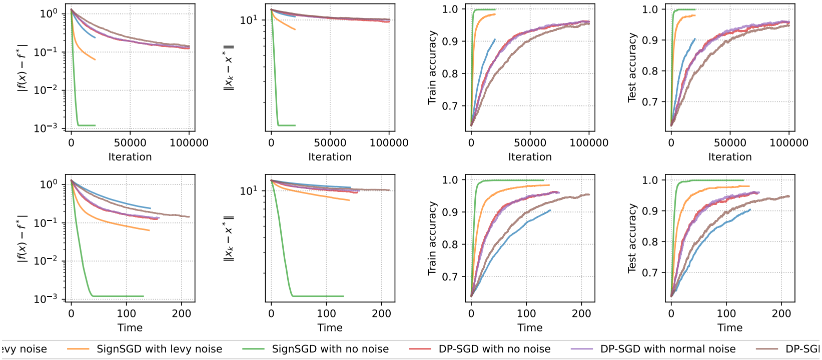
While according to the advanced composition theorem and conversion from Rényiprivacy to (ε, δ)-privacy, ε_R must satisfy:

$$\varepsilon_R \leq \varepsilon / T - \frac{\log 1/\delta}{T(\alpha - 1)}$$

.

Grid Search to find minimal σ





dp-sign SGD converges very slowly, but it depends on a dataset.
Lower q leads to much lower σ and better convergence.

Conclusion

dp-sign SGD

- with Poisson sampling $q = 1/n$
- is $(10, 1/n^{1.1})$ differentially-private
- empirically converges on logistic regression problem even with heavy-tailed noise

Now we have to provide theoretical guarantees of convergence.