Faster Rates of Differentially Private stochastic Convex Optimization

Jinyan Su
UESTC

Lijie Hu
KAUST

Di Wang
KAUST

Work done at KAUST

Tsybakov Noise Condition (TNC)

Definition: (θ, λ)-TNC

- $\theta > 1, \lambda > 0$
- Convex function $F(\cdot)$
- $w^* = \arg\min_{w \in \mathscr{C}} F(w)$
- $F(w) - F(w^*) \geq \lambda ||w - w^*||_2^\theta, \forall w \in \mathscr{C}$

Tsybakov Noise Condition (TNC)

Examples:

- λ-strongly convex: $(2, \frac{\lambda}{2})$-TNC
- Weak strong convexity
- Error Bound(EB)
- Polyak-Lojasiewicz (PL) functions

DP-SCO

Stochastic Convex Optimization (SCO):

- Unknown distribution $\mathscr{P}$ over data universe $\mathscr{X}$
- Dataset $S = \{x_1, x_2, \cdots, x_n\}$
- Convex set $\mathscr{C} \subseteq \mathbb{R}^d$
- Convex loss function $f : \mathscr{C} \times \mathscr{X} \to \mathbb{R}$
- Excess population loss: $\mathbb{E}_{x \sim \mathscr{P}} [f(\hat{w}, x)] - \min_{w \in \mathscr{C}} \mathbb{E}_{x \sim \mathscr{P}} [f(w, x)]$

DP-SCO

Differential Privacy (DP)

- Neighboring Datasets $S, S^{'} \subseteq \mathscr{X}$
- Random algorithm: $\mathscr{A}$ with output space E
- $\mathscr{A}$ is (ϵ, δ)-DP if $Pr(\mathscr{A}(S) \in E) \leq e^{\epsilon} Pr(\mathscr{A}(S^{'}) \in E) + \delta$

Previous results of DP-SCO

- Convex function: $O\left(\frac{1}{\sqrt{n}} + \frac{\sqrt{d\log(1/\delta)}}{n\epsilon}\right)$
- Strongly convex function: $O\left(\frac{1}{n} + \frac{d\log(1/\delta)}{n^2\epsilon^2}\right)$

What makes strongly convex result better?

$$F(w) - F(w^*) \geq \frac{\lambda}{2}||w - w^*||_2^2$$

TNC：

$$F(w) - F(w^*) \geq \lambda||w - w^*||_2^\theta, \forall w \in \mathscr{C}$$

Contribution 1

Algorithms for TNC and their Excess population risk

1. Private Stochastic Approximation

$\theta \geq 2$ & Projection into $\mathscr{C} \cap \mathbb{B}(\hat{w}_{k-1}, R_{k-1})$

2. Private Stochastic Approximation-II

Upper bound =

$$\left(\frac{\sqrt{\log n}}{\sqrt{n}} + \frac{\sqrt{d}\log n}{n\epsilon}\right)^{\frac{\theta}{\theta-1}}$$

Known θ

3. Iterated Phased-SGD

$\theta \geq \bar{\theta} > 1$

Upper bound =

$$\left(\frac{1}{\sqrt{n}} + \frac{\sqrt{d}}{n\epsilon}\right)^{\frac{\theta}{\theta-1}}$$

Lower bound:

$$\Omega\left(\left(\frac{\sqrt{d}}{n\epsilon}\right)^{\frac{\theta}{\theta-1}}\right)$$

$\theta \geq 2$

Comparison to [Asi et al., 2021b]

- Different proofs
- [Asi et al., 2021b] Projection into a ball and a set
- Alternative algorithms

Contribution 2

Improved rate for strongly convex loss

- Assumptions: Non-negative smooth loss & Small optimal value $F(w^*)$
- $n > \kappa^\tau, \kappa = \frac{\beta}{\lambda}$
- Faster-DPSGD-SC: $O\left(\frac{1}{n^\tau} + \frac{d}{n^2\epsilon^2}\right), \tau > 1$

SC: $O\left(\frac{1}{n} + \frac{d}{n^2\epsilon^2}\right)$

Private Stochastic Approximation

Phased-SGD($w_0, \mathscr{C}$) [Feldman et al., 2020]

Initial point w_0

For $i = 1, \cdots, m$ do

Compute the average iterate of PSGD: $\bar{w}_i$

(Projection into $\mathscr{C}$)

Add noise: $w_i = \bar{w}_i + \xi_i$

End

(Sketch):

Partite the dataset equally into m parts

For $k = 1, \cdots, m$ do

$\hat{w}_k$ = Phased-SGD($\hat{w}_{k-1}, \mathscr{C} \cap \mathbb{B}(\hat{w}_{k-1}, R_{k-1})$)

End

Main technique:

A shrinking ball centered at the output of the last iterate

Private Stochastic Approximation-II

(Sketch):

Partite the dataset equally into m parts

For $k = 1, \cdots, m$ do

Use Phased-SGD-SC [Feldman et al., 2020] to solve

$$w^k = \arg\min F(w) + \frac{1}{2\gamma_k} ||w - w_{k-1}||_2^2$$

End

Main technique:

1. Additional strongly convex regularization
2. Solve with strongly convex version of phased-SGD

Iterated Phased-SGD

Main idea: reduction to the convex case

(Sketch):

Partite dataset into subsets $\{S_1, \cdots, S_k\}$

For $t = 1, \cdots, k$ **do**

w_t = Phased-SGD(w_{t-1} , $\mathscr{C}$) using S_t

(Initialized at the output of previous phase)

End

Phased-SGD($w_0, \mathscr{C}$) [Feldman et al., 2020]

Initial point w_0

For $i = 1, \cdots, m$ **do**

Compute the average iterate of PSGD: $\bar{w}_i$

(Projection into $\mathscr{C}$)

Add noise: $w_i = \bar{w}_i + \xi_i$

End

Faster-DPSGD- SC

(Sketch):

Split dataset into 2 equal subsets $|S_1| = |S_2|$

Perform the Iterated Phased-SGD on S_1 to return $\hat{w}$

Perform Epoch-DP-SGD on S_2

Open problems

- Close the gap of $O\left(\frac{1}{n^{\frac{\theta}{2(\theta-1)}}}\right)$ between upper and lower bound
- Faster rate for other special class of functions, e.g., exponential concave

Thank you!