Open problem: Nearly Linear-Time algorithm for DP-ERM

Yu-Xiang Wang

A quick writeup here:
https://shorturl.at/nMRZ4

Differentially Private Empirical Risk Minimization

- Originated in [Chaudhuri, Monteleoni, Sarwate, 2011]
- Data: $(x_1, y_1), ..., (x_n, y_n) \in \mathcal{X} \times \mathcal{Y} = \mathcal{Z}$
- Problem: Approximately solve Convex ERM

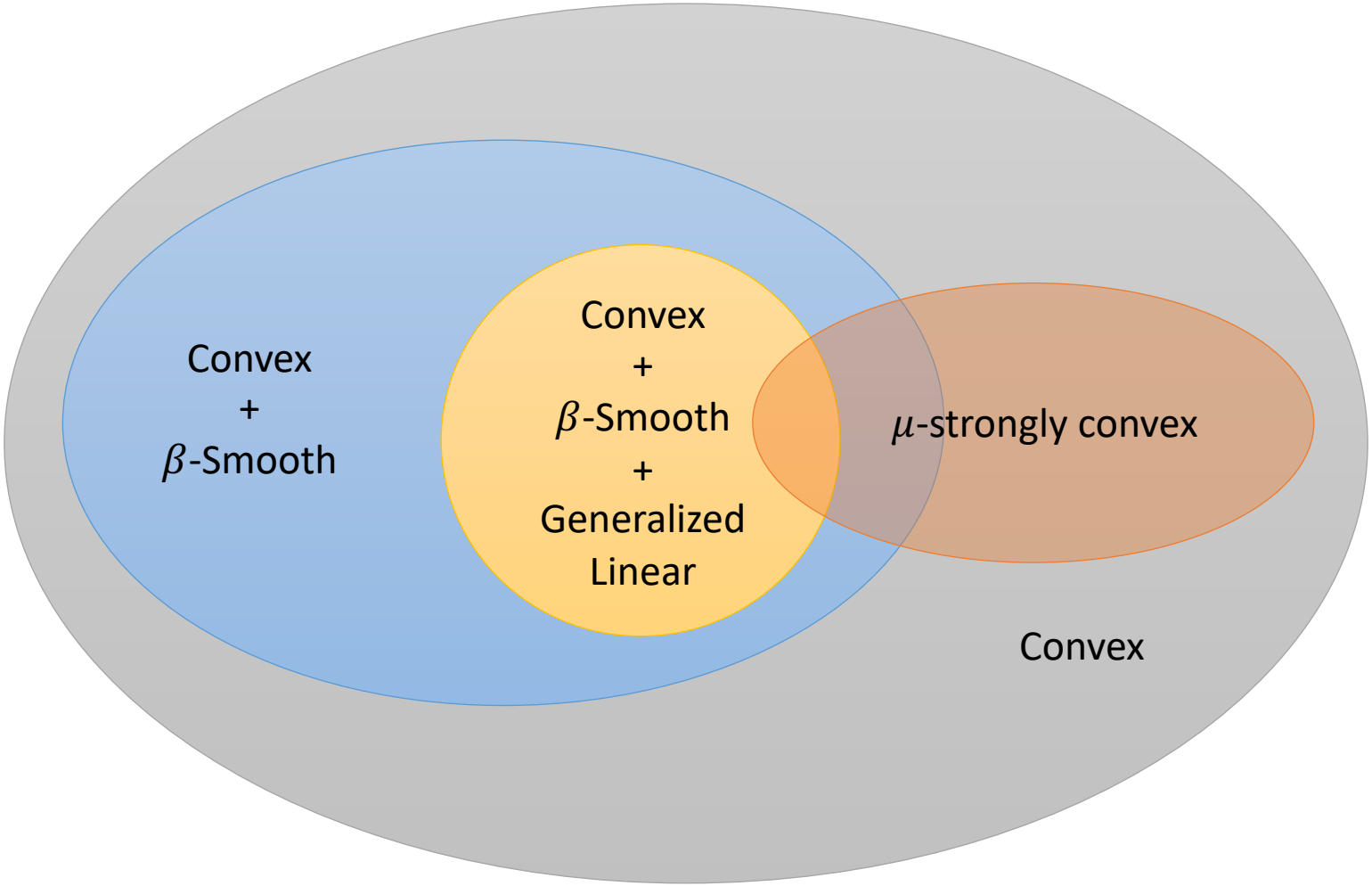
$$\min_{\theta \in \Theta \subset \mathbb{R}^d} \sum_{i=1}^{n} \ell(\theta, (x_i, y_i))$$

  - Subject to an (ϵ, δ)-differential privacy constraint.
- Utility metric: Excess Empirical Risk

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \ell_i(\hat{\theta})\right] - \frac{1}{n}\sum_{i=1}^{n} \ell_i(\theta^*)$$

Hierarchy of DP-ERM problems

Assume $\ell(\theta, (x, y))$ is L-Lipschitz in θ, and …



The information-theoretic limit is known since [KST-12][BST-14]

	Lipschitz + convex	Lipschitz + Smooth + convex	Smooth + Lipschitz + convex + GLM
ObjPert	Not applicable	$\frac{dL\|\theta^*\|\sqrt{\log(\frac{1}{\delta})}}{n\epsilon}$	$\frac{\sqrt{d}L\|\theta^*\|\sqrt{\log(\frac{1}{\delta})}}{n\epsilon}$
NoisyGD	$\frac{\sqrt{d}L\|\theta^*\|\sqrt{\log(\frac{1}{\delta})}}{n\epsilon}$	$\frac{\sqrt{d}L\|\theta^*\|\sqrt{\log(\frac{1}{\delta})}}{n\epsilon}$	$\frac{\sqrt{d}L\|\theta^*\|\sqrt{\log(\frac{1}{\delta})}}{n\epsilon}$

Lower order terms and dependence on β hidden.

	Lipschitz + Strongly Convex
NoisyGD	$\frac{dL^2\log(1/\delta)}{\mu n^2\epsilon^2}$

But surprisingly, there is no known algorithm that achieves the statistical limit for DP-ERM while runs in nearly linear time.

- Even if we assume smoothness.
- Even if we assume strong convexity.

Computational Complexity

- in terms of the number of incremental gradient calls to achieve information theoretic limit up to a constant

	Lipschitz + Smooth + Convex	Lipschitz + Convex	Lipschitz + Strongly Convex
NoisyGD	$\frac{n^2\epsilon}{d^{1/2}}$	$\frac{n^3\epsilon^2}{d}$	$\frac{n^3\epsilon^2}{d}$
NoisySGD [BST'14]	$\frac{n^{3/2}\epsilon}{d^{1/4}} + \frac{n^2\epsilon^2}{d}$	$\frac{n^2\epsilon^{3/2}}{d^{1/2}} + \frac{n^2\epsilon^2}{d}$	$\frac{n^2\epsilon^{3/2}}{d^{1/2}} + \frac{n^2\epsilon^2}{d}$
Posterior Sampling	[LST'22] $n + n^{3/2}\epsilon$	[GLL'22] $\frac{n^2\epsilon^2}{d}$	[GLL'22] $\frac{n^2\epsilon^2}{d}$

Open problem: Can we obtain nearly linear time algorithm?

Recent progress: Revisiting ObjPert

For Convex, Smooth + GLM loss case, there is an O(n \log n) time algorithm that achieves the optimal rate.

The algorithm is a combination of **Objective Perturbation**, **Approximate Minima Perturbation**, and **SVRG** with a carefully chosen regularization parameter. [Redberg, Honkela, W., 2023]

If one drops the GLM condition, the same algorithm runs in O(n \log n) but incurs additional dimension-dependence. [Redberg, Honkela, W., 2023]

Recent progress: Posterior sampling via DP-SGLD

[Chourasia, Ye, Shokri, 2021]
[Ryffel, Bach, Pointcheval, 2022]

- Strongly convex + smooth case

	Excess Empirical Risk	Computation
DP-Langevin Dynamics [CYS-21]	$\frac{\beta}{\mu}\frac{dL^2\log(1/\delta)}{\mu n^2\epsilon^2}$	$\frac{\beta}{\mu}\cdot n\log n$
DP-SGLD [RBP-22, Thm 4.2]	$\frac{\beta}{\mu}\frac{dL^2\log(1/\delta)}{\mu n^2\epsilon^2}+\frac{L^2}{\mu n^\alpha}$	$\frac{\beta}{\mu}\cdot n^\alpha\log n$ for $0<\alpha<1$
DP-SGLD [RBP-22, Thm 5.1]	$\frac{dL^2\log(1/\delta)}{\mu n^2\epsilon^2}$	At least $\frac{\beta}{\mu}\cdot\frac{n^2\epsilon^2}{d}$

Recent progress: Noisy Newton Method [Avella-Medina, Bradshaw, Loh, 2022]

- For the strongly-convex + smooth + self-concordance loss functions

	Excess Empirical Risk	Computation
Noisy-Newton Method [ABL22]	$\log\log n \cdot \frac{dL^2 \log(1/\delta)}{\mu n^2 \epsilon^2}$	$\log\log n \cdot n$

Other possible leads

- For the Convex+Smooth case
  - New analysis of Noisy-SGD by [Altschuler and Talwar, 2022] , however does not lead to improvements unless going to the slow SGD regime, at least quadratic time.
  - Other MCMC methods for implementing Posterior Sampling mechanism.

Wisdom from optimization theory

[Woodworth and Srebro, 2016] [Zhou and Gu, 2019] [Allen-Zhu, 2017]

- Smooth and convex finite sum problems

$$n + \sqrt{n\beta/\mathrm{SubOpt}} \quad \Rightarrow \quad O(n + n\sqrt{\epsilon})$$

$$\mathrm{SubOpt} = O(1/n\epsilon)$$

- Smooth and strongly convex finite sum problems

$$n + \sqrt{\frac{n\beta}{\mu}} \log\left(\frac{1}{\mathrm{SubOpt}}\right) \quad \Rightarrow \quad \tilde{O}(n) \text{ for all } \mu > \frac{1}{n\epsilon}$$

$$\mathrm{SubOpt} = O(1/\mu n^2 \epsilon^2)$$

One possible way to resolve the open problem

- [Lee, Shen, Tian, 2020] Finite-sum Log-concave sampling --- convergence takes the following number of steps

$$n + \frac{\beta}{\mu}\sqrt{n}\log\left(\frac{1}{\text{TV-dist}}\right) \Rightarrow \tilde{O}(n^{3/2}) \quad \text{when } \mu = 1/n\epsilon$$

- Can we derive *"Accelerated" version* of *finite sum log-concave sampling*, improve from

$$n + \sqrt{\frac{\beta}{\mu}}\sqrt{n}\log\left(\frac{1}{\text{TV-dist}}\right)$$

Summary

- Surprisingly challenging to get optimal privacy-utility tradeoff in subquadratic time

- Nearly linear time in specific settings:
  - Convex, Smooth GLM --- ObjPert-AMP with SVRG [Redberg, Honkela, W., 2023]
  - Strongly convex, Smooth, with condition number $< \sqrt{n}$ [Lee, Shen, Tian, 2020]

- Open in all other cases!