CSC/DSCC 574 Continuous Algorithms for Optimization and Sampling Lecture 12

Revisiting Differential Privacy

Lecturer: Jiaming Liang *April 23, 2024*

1 Differential privacy

Definition 1 ((ε, δ)-DP)**.** *A randomized mechanism $\mathcal{M}$ is (ε, δ)-differentially private if for any neighboring databases $\mathcal{D}, \mathcal{D}'$ and any subset $S \subseteq O$ (output space), one has*

$$\mathbb{P}[\mathcal{M}(\mathcal{D}) \in S] \leq e^{\varepsilon} \mathbb{P}\left[\mathcal{M}\left(\mathcal{D}'\right) \in S\right] + \delta.$$

We say $\mathcal{D}$ and $\mathcal{D}'$ are neighboring databases if they agree on all the user inputs except for a single user's input.

For $\delta = 0$, the ε-DP condition can be written as

$$\frac{1}{e^{\varepsilon}} \mathbb{P}\left[\mathcal{M}\left(\mathcal{D}'\right) \in S\right] \leq \mathbb{P}[\mathcal{M}(\mathcal{D}) \in S] \leq e^{\varepsilon} \mathbb{P}\left[\mathcal{M}\left(\mathcal{D}'\right) \in S\right].$$

A DP algorithm $\mathcal{M}$ usually satisfies a collection of (ε, δ)-DP guarantees for each ε, i.e., for each $\varepsilon \geq 0$, there exists a smallest δ for which $\mathcal{M}$ is (ε, δ)-DP. By collecting all of them together, we can form the privacy curve or privacy profile that fully characterizes the privacy of a DP algorithm.

Definition 2 (Privacy Curve)**.** *Given two random variables X and Y supported on some set Ω, define the privacy curve $\delta(X \| Y) : \mathbb{R}_{\geq 0} \to [0, 1]$ as follows,*

$$\delta(X \| Y)(\varepsilon) = \sup_{S \subset \Omega} \Pr[Y \in S] - e^{\varepsilon} \Pr[X \in S].$$

We say a differentially private mechanism $\mathcal{M}$ has privacy curve $\delta : \mathbb{R}_{\geq 0} \to [0, 1]$ if for every $\varepsilon \geq 0, \mathcal{M}$ is $(\varepsilon, \delta(\varepsilon))$-differentially private, i.e., $\delta\left(\mathcal{M}(\mathcal{D}) \| \mathcal{M}\left(\mathcal{D}'\right)\right)(\varepsilon) \leq \delta(\varepsilon)$ for all neighbouring databases $\mathcal{D}$ and $\mathcal{D}'$.

We will also need the notion of tradeoff function, which is an equivalent way to describe the privacy curve $\delta(P \| Q)$.

Definition 3 (Tradeoff function)**.** *Given two (continuous) distributions P and Q, we define the tradeoff function $T(P \| Q) : [0, 1] \to [0, 1]$ as*

$$T(P \| Q)(z) = \inf_{S : P(S) = 1 - z} Q(S).$$

The tradeoff function $T(P \| Q)$ and the privacy curve $\delta(P \| Q)$ are related via convex duality. Therefore to compare privacy curves, it is enough to compare tradeoff curves.

Lemma 1. *We have*

$$\delta(P \| Q) \leq \delta\left(P' \| Q'\right) \text{ iff } T(P \| Q) \geq T\left(P' \| Q'\right).$$

2 Private convex optimization

Recall stochastic optimization is

$$\min_{x \in Q} \{ f(x) = \mathbb{E}_{\xi}[F(x;\xi)] \},$$

and its sample average approximation is

$$f(x;\mathcal{D}) = \frac{1}{n}\sum_{i=1}^{n} F(x;\xi_i),$$

where $\mathcal{D} = \{\xi_1, \ldots, \xi_n\}$ is a database.

Here $f(x;\mathcal{D})$ can be understood as the negative utility function $-u(\mathcal{D};s)$, where $x = s$ is the output of a certain mechanism. We assume $F(\cdot;\xi)$ is convex and M-Lipschitz continuous, and Q has a diameter $D > 0$.

We want to output a solution x^{priv} using a differentially private mechanism $\mathcal{M}$ such that we minimize the excess empirical risk

$$\mathbb{E}_{\mathcal{M}}\left[f\left(x^{\text{priv}};\mathcal{D}\right)\right] - f\left(x_*;\mathcal{D}\right),$$

where $x_* \in Q$ is the minimizer of $f(x;\mathcal{D})$.

In the literature, it is shown that EM achieves the optimal excess empirical risk $\mathcal{O}\left(\frac{MDd}{n\varepsilon}\right)$ under ε-DP. On the other hand, it has also been shown that noisy gradient descent achieves an excess empirical risk of

$$\mathcal{O}\left(\frac{MD\sqrt{d\log\frac{1}{\delta}}}{n\varepsilon}\right)$$

under (ε,δ)-DP, which is also shown to be optimal.

Note that the second bound only loses a bit in privacy (δ) but reduces the dependence of d in the excess empirical risk from d to $\sqrt{d}$. It is natural to ask the question whether we can obtain the optimal empirical risk under (ε,δ)-DP using EM. The answer is affirmative, but we need to introduce a modified version of EM, that is the regularized exponential mechanism,

$$x^{\text{priv}} \sim \exp\left(-k\left[f(x;\mathcal{D}) + \frac{\mu}{2}\|x\|_2^2\right]\right).$$

With a suitable choice of μ and k, we recover the optimal excess risk under (ε,δ)-DP.

EM is the task of sampling and the regularized EM is an instance of the restricted Gaussian oracle that we have studied in proximal sampling. Since we have studied the non-asymptotic convergence of sampling algorithms, we are ready to establish the excess empirical risk using the regularized EM.

3 Analysis

Theorem 1. *Given convex set $\mathcal{K} \subseteq \mathbb{R}^d$ and μ-strongly convex functions F and $\tilde{F}$ over $\mathcal{K}$. Let P and Q be distributions over $\mathcal{K}$ such that $P(x) \propto \exp(-F(x))$ and $Q(x) \propto \exp(-\tilde{F}(x))$. If $\tilde{F} - F$ is G-Lipschitz over $\mathcal{K}$, then for all $\varepsilon > 0$, we have*

$$\delta(P\|Q)(\varepsilon) \leq \delta\left(\mathcal{N}(0,1)\|\mathcal{N}\left(\frac{G}{\sqrt{\mu}},1\right)\right)(\varepsilon)$$
$$T(P\|Q)(z) \geq T\left(\mathcal{N}(0,1)\|\mathcal{N}\left(\frac{G}{\sqrt{\mu}},1\right)\right)(z).$$

This proves that the privacy curve for distinguishing between P and Q is upper bounded by the privacy curve of a Gaussian mechanism with sensitivity $G/\sqrt{\mu}$ and noise scale 1 .

Theorem 2 (Kalai and Vempala)**.** *Let $f(x) = c^T x$, where c is a unit vector, and let $\mathcal{K} \subseteq \mathbb{R}^d$ be a convex body. Then, for any $t > 0$, we have*

$$\mathop{\mathbb{E}}_{X \sim P_{\frac{1}{t}f}}[f(X)] - \min_{x \in \mathcal{K}} f(x) \leq nt.$$

Extension to an arbitrary convex function f.

Lemma 2 (Utility Guarantee)**.** *Suppose $k > 0$ and F is a convex function over a convex body $\mathcal{K} \subseteq \mathbb{R}^d$. For the distribution $\nu(x) \propto \exp(-kf(x))$, we have*

$$\mathop{\mathbb{E}}_{\nu}[f(x)] \leq \min_{\mathcal{K}} f(x) + \frac{d}{k}.$$

Proof. Define

$$E_{\mathcal{K}} := \mathop{\mathbb{E}}_{X \sim P_{\frac{1}{t}f}}[f(X)] = \frac{\int_{\mathcal{K}} f(x) e^{-f(x)/t} dx}{\int_{\mathcal{K}} e^{-f(x)/t} dx}.$$

It is clear that

$$\min_{x \in \mathcal{K}} f(x) \leq E_{\mathcal{K}}.$$

Define the set

$$\hat{\mathcal{K}} := \left\{(x, x_{n+1}) \in \mathbb{R}^{n+1} : x \in \mathcal{K}, f(x) \leq x_{n+1} \leq E_{\mathcal{K}}\right\}.$$

Then $\hat{\mathcal{K}}$ is a convex body, and we have

$$\min_{x \in \mathcal{K}} f(x) = \min_{(x, x_{n+1}) \in \hat{\mathcal{K}}} x_{n+1}.$$

Accordingly, define the parameter

$$E_{\hat{\mathcal{K}}} := \frac{\int_{\hat{\mathcal{K}}} x_{n+1} e^{-x_{n+1}/t} dx_{n+1} dx}{\int_{\hat{\mathcal{K}}} e^{-x_{n+1}/t} dx_{n+1} dx}.$$

Nextf, we show that

$$E_{\hat{\mathcal{K}}} = E_{\mathcal{K}} + t. \tag{1}$$

To this end, set $E_{\mathcal{K}} = N_{\mathcal{K}}/D_{\mathcal{K}}$ and $E_{\hat{\mathcal{K}}} = N_{\hat{\mathcal{K}}}/D_{\hat{\mathcal{K}}}$, where we define

$$\begin{aligned} N_{\mathcal{K}} &:= \textstyle\int_{\mathcal{K}} f(x) e^{-f(x)/t} dx, & D_{\mathcal{K}} &:= \textstyle\int_{\mathcal{K}} e^{-f(x)/t} dx, \\ N_{\hat{\mathcal{K}}} &:= \textstyle\int_{\hat{\mathcal{K}}} x_{n+1} e^{-x_{n+1}/t} dx_{n+1} dx, & D_{\hat{\mathcal{K}}} &:= \textstyle\int_{\hat{\mathcal{K}}} e^{-x_{n+1}/t} dx_{n+1} dx. \end{aligned}$$

We work out the parameters $N_{\hat{\mathcal{K}}}$ and $D_{\hat{\mathcal{K}}}$ (taking integrations by part):

$$D_{\hat{\mathcal{K}}} = \int_{\mathcal{K}} \left(\int_{f(x)}^{E_{\mathcal{K}}} e^{-x_{n+1}/t} dx_{n+1} \right) dx = \int_{\mathcal{K}} \left(t e^{-f(x)/t} - t e^{-E_{\mathcal{K}}/t} \right) dx = t D_{\mathcal{K}} - t e^{-E_{\mathcal{K}}/t} \operatorname{vol}(\mathcal{K}),$$

$$\begin{aligned} N_{\hat{\mathcal{K}}} &= \int_{\mathcal{K}} \left(\int_{f(x)}^{E_{\mathcal{K}}} x_{n+1} e^{-x_{n+1}/t} dx_{n+1} \right) dx = \int_{\mathcal{K}} \left(-t E_{\mathcal{K}} e^{-E_{\mathcal{K}}/t} + t f(x) e^{-f(x)/t} + t \int_{f(x)}^{E_{\mathcal{K}}} e^{-x_{n+1}/t} dx_{n+1} \right) dx \\ &= -t E_{\mathcal{K}} e^{-E_{\mathcal{K}}/t} \operatorname{vol}(\mathcal{K}) + t N_{\mathcal{K}} + t D_{\hat{\mathcal{K}}}. \end{aligned}$$

Then, using the fact that $E_{\mathcal{K}} = N_{\mathcal{K}}/D_{\mathcal{K}}$, we obtain

$$\frac{N_{\hat{\mathcal{K}}}}{D_{\hat{\mathcal{K}}}} = t + \frac{N_{\mathcal{K}} - E_{\mathcal{K}} e^{-E_{\mathcal{K}}/t} \operatorname{vol}(\mathcal{K})}{D_{\mathcal{K}} - e^{-E_{\mathcal{K}}/t} \operatorname{vol}(\mathcal{K})} = t + \frac{N_{\mathcal{K}}}{D_{\mathcal{K}}},$$

which proves relation (1). Now we are ready to prove the lemma. Indeed, using Theorem 2 applied to $\hat{\mathcal{K}}$ and the linear function x_{n+1}, we get

$$\mathop{\mathbb{E}}_{X \sim P_{\frac{1}{t} f}} [f(X)] - \min_{x \in \mathcal{K}} f(x) = E_{\mathcal{K}} - \min_{x \in \mathcal{K}} f(x) = \left(E_{\hat{\mathcal{K}}} - \min_{(x, x_{n+1}) \in \hat{\mathcal{K}}} x_{n+1} \right) + \left(E_{\mathcal{K}} - E_{\hat{\mathcal{K}}} \right) \leq t(n+1) - t = tn.$$

$\square$

Theorem 3 (Main). *Let $\varepsilon > 0, \mathcal{K} \subseteq \mathbb{R}^d$ be a convex set of diameter D and $\{F(\cdot, \xi)\}_{\xi \in \mathcal{D}}$ be a family of G-Lipschitz functions over $\mathcal{K}$. For any data-set $\mathcal{D}$ and $k > 0$, sampling $x^{(priv)}$ with probability proportional to* $\exp\left(-k\left(f(x; \mathcal{D}) + \mu \|x\|_2^2/2\right)\right)$ *is $(\varepsilon, \delta(\varepsilon))$-differentially private, where*

$$\delta(\varepsilon) \leq \delta \left(\mathcal{N}(0,1) \| \mathcal{N}\left(\frac{2G\sqrt{k}}{n\sqrt{\mu}}, 1 \right) \right)(\varepsilon).$$

The excess empirical risk is bounded by $\frac{d}{k} + \frac{\mu D^2}{2}$.