Composition Theorem for streaming CW-pricacy

1 Characterize Privacy as Regions

In this section, we show how to characterize Differential Privacy (DP) and CW-Privacy (CW-P) in terms of convex regions, and how to compute the (ϵ, δ) values for each mechanism in terms of tangent lines of such convex regions. We focus on the context of finite discrete states for the simplicity of analysis.

1.1 Case of DP

Let $\mathcal{X}$ be the set of all databases. Let $\mathcal{Y}$ be the set of all out comes of mechanism M. M is a probability measure from $\mathcal{X}$ to $\mathcal{Y}$. For simplicity of analysis, assume $|\mathcal{X}| = n$ and $|\mathcal{Y}| = m$. Then, M corresponds to an $n \times m$ Markov matrix $M = [M(x_1), \ldots, M(x_n)]^\top$.

Now, it is easy to see the following is an equivalent definition of DP.

Theorem 1. *For any $\epsilon \geq 0$ and $\delta \in [0, 1]$, a mechanism M is (ϵ, δ)-differentially private if and only if the following conditions are satisfied for all pairs of neighboring databases x and x', and all region $S \subseteq \mathcal{Y}$:*

$$Pr[M(x) \in S] + e^\epsilon Pr[M(x') \in \bar{S}] \geq 1 - \delta, \quad \textit{and}$$
$$e^\epsilon Pr[M(x) \in S] + Pr[M(x') \in \bar{S}] \geq 1 - \delta.$$

This gives a graphical representation (region) of DP:

$$R(\epsilon, \delta) = \{(p_x, p_y) \,|\, p_x + e^\epsilon p_y \geq 1 - \delta, e^\epsilon p_x + p_y \geq 1 - \delta\}.$$

For any two databases x and x', define

$$R(M, x, x') = convex\{(Pr[M(x) \in S], Pr[M(x') \in \bar{S}]) \,|\, \text{for all } S \subseteq \mathcal{Y}\}$$

$R(M, x, x')$ has the following equivalent form.

$$R(M, x, x') = \{(M(x) \cdot \alpha, M(x') \cdot \beta) \,|\, 0 \leq \alpha_i, \beta_i \leq 1,\ \alpha + \beta = \mathbf{1^m}\},$$

where $\cdot$ denote the dot production of vectors.

Definition 1. *For any mechanism M, we define its privacy region $R(M) = \bigcup_{(x,x')} R(M, x, x')$, where (x, x') is a pair of neighboring databases.*

Immediately, it should not hard to see the following theorem.

Theorem 2. *M is (ϵ, δ)-differentially private iff $R(M) \subseteq R(\epsilon, \delta)$.*

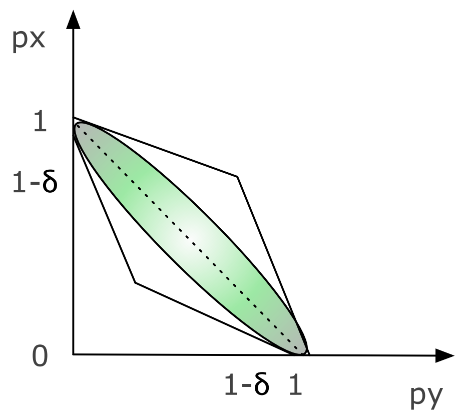

Figure 1: Every tangent line of the privacy region forms a pair of (ϵ, δ).

As it is shown in Figure 1, every tangent line of the privacy region corresponds to a pair of (ϵ, δ). In general, the privacy region $R(M, x, x')$ of any mechanism M can be represented by the intersections of all such regions $\{R(\epsilon_i, \delta_i)\}$, which is completely described by the set of slopes and shifts $\{(\epsilon_i, \delta_i)\}$.

1. For the set slopes, let $\mathcal{E} = \{0 \leq \epsilon_i < \infty \,|\, Pr[M(x) = y] = e^{\epsilon_i} Pr[M(x') = y] \text{ for some } y \in \mathcal{Y}\}$

2. For each ϵ_i, $\delta_i = \max_{S \subseteq \mathcal{Y}} \{\Sigma_{y \in S} Pr[M(x) = y] - e^{\epsilon_i} \Sigma_{y \in S} Pr[M(x) = y]\}$

1.2 Case of CW-P

In the case of CW-P, we use X to denote a database variable that follows some distribution D. Let the pmf of D be $f_D = [f_{x_1}, \ldots, f_{x_n}]$. Let $alt(X)$ denote a scrubbed version of X. Assume $alt(X)$ follows some distribution D' with pmf f'_D.

It is easy to see that CW-P has the following equivalent definition.

Theorem 3. *For any $\epsilon \geq 0$ and $\delta \in [0,1]$, a mechanism M is (ϵ, δ)-differentially private if and only if the following conditions are satisfied for all distributions on D on (X, Z), all $(priv, alt)$ pairs, and all region $S \subseteq \mathcal{Y}$:*

$$Pr[M(X) \in S \,|\, priv(X), Z] + e^{\epsilon} Pr[M(alt(X)) \in \bar{S} \,|\, priv(X), Z] \geq 1 - \delta, \quad and$$

$$e^{\epsilon} Pr[M(X) \in S \,|\, priv(X), Z] + Pr[M(alt(X)) \in \bar{S} \,|\, priv(X), Z] \geq 1 - \delta,$$

where $M(X) = f_D M$ and $M(alt(X)) = f'_D M$.

Notice mechanism $M' = [f_D M, f'_D M]^\top$ can be seen as a DP version of M. Hence, CW-P also has form of privacy regions, and everything else described above follows.

2 Composition Theorem of CW-P in the streaming setting