

Theory of Differential Privacy

Lecture 5:

Advanced Composition

EN.601.438/638 Fall 2026

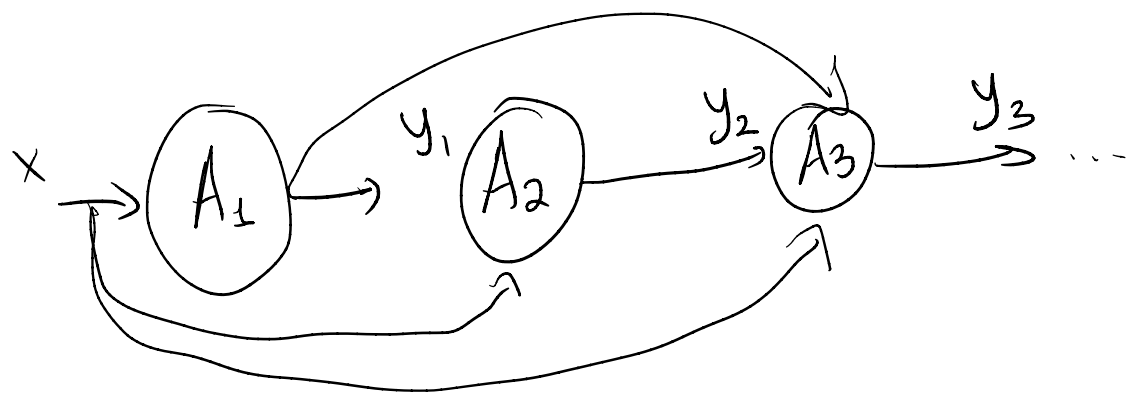
Lydia Zakynthinou
Johns Hopkins University

Sept 16, 2026

(k-fold) Adaptive Composition

$A_1, \dots, A_k$

$$Y_1 \sim A_1(x), \dots, Y_j | (Y_{<j} = y_{<j}) \sim A_j(x; y_{<j}) \quad j=2, \dots, k$$



Transcript : $A(x) = (Y_1, \dots, Y_k)$

Assumption (A1)

For every fixed j , every fixed history $h = y_{<j}$, every neighboring $x \sim x'$

every event E :

$$\Pr[A_j(x; h) \in E] \leq e^\epsilon \Pr[A_j(x'; h) \in E] + \delta$$

i.e.:

$\forall h : z \mapsto A_j(z; h)$ is (ϵ, δ) -DP.

Advanced Composition

Let $\varepsilon \geq 0$, $\delta \in (0, 1)$, $\delta' \in (0, 1)$ s.t. $k\delta + \delta' < 1$. Under (A1), A is $(\tilde{\varepsilon}, \tilde{\delta})$ -DP, where:

$$\begin{aligned}\tilde{\varepsilon} &= \varepsilon \sqrt{2k \log(1/\delta')} + k \cdot \varepsilon \cdot \frac{\overbrace{e^\varepsilon - 1}}{e^\varepsilon + 1} \\ &\leq \varepsilon \sqrt{2k \log(1/\delta')} + k \varepsilon^2 / 2.\end{aligned}$$

$$\tilde{\delta} = k\delta + \delta'$$

$$\approx \quad \tilde{\varepsilon} = \Theta\left(\varepsilon \sqrt{k \log(1/\delta')}\right).$$

$$\tilde{\delta} = \Theta(k\delta)$$

Privacy loss of transcript

- $P_{X,j}(y_j | y_{<j}) = \Pr[A_j(x; y_{<j}) = y_j]$

- $P_{X',j}$

$$P_X(y_{1:k}) = \prod_{j=1}^k P_{X,j}(y_j | y_{<j})$$

$$P_{X'}(y_{1:k}) = \prod_{j=1}^k P_{X',j}(y_j | y_{<j})$$

for fixed output $y_{1:k}$:

$$L_{X \rightarrow X'}(y_{1:k}) = \log \frac{P_{X,j}(y_{1:k})}{P_{X',j}(y_{1:k})} = \sum_{j=1}^k \log \frac{P_{X,j}(y_j | y_{<j})}{P_{X',j}(y_j | y_{<j})}$$

(let $y_{1:k} \sim A(x)$.)

$$\text{PLRV: } L_{X \rightarrow X'}(y_{1:k}) = \sum_{j=1}^k \log \frac{P_{X,j}(y_j | y_{<j})}{P_{X',j}(y_j | y_{<j})}$$

basic composition:

$$\Pr_{Y_j} [e_{X \rightarrow X'}(y_j | y_{<j}) > \epsilon | y_{<j} = h] \leq \delta \quad \forall j, h$$

$$\Pr_{Y_{1:k}} \left[\bigcup_{i=1}^k \{e_{X \rightarrow X'}(y_i | y_{<i}) > \epsilon\} \right]$$

$$\leq \sum_{j=1}^k \Pr_{Y_{1:k}} [e_{X \rightarrow X'}(y_j | y_{<j}) > \epsilon]$$

$$= \sum_{j=1}^k \mathbb{E}_{h \sim Y_{<j}} \left[\Pr_{Y_j} [e_{X \rightarrow X'}(y_j | y_{<j}) > \epsilon | y_{<j} = h] \right]$$

$$\leq \sum_{j=1}^k \delta \leq k\delta.$$

So w.p. $1 - k\delta$, all $e_{X \rightarrow X'}(y_j | y_{<j}) \leq \epsilon$

$$\Rightarrow L_{X \rightarrow X'}(y_{1:k}) \leq k\epsilon$$

Warm-up: Randomized Response PLRV

$$P = \frac{e^{\varepsilon}}{1 + e^{\varepsilon}}, \quad \text{output } x$$

		values	
		0	1
Random variable of output under input 0	$\rightarrow R_0$	p	$1-p$
"	$\rightarrow R_1$	$1-p$	p

Draw $Z \sim P_0$: $X := \log \frac{Pr[R_0 = Z]}{Pr[R_1 = Z]}$: PLRV of RR_E .

then
$$X = \begin{cases} +\varepsilon & , \text{w.p. } p \text{ (when } Z=0) \\ -\varepsilon & , \text{w.p. } 1-p \text{ (when } Z=1) \end{cases}$$

$$\mathbb{E}[X] = \varepsilon \cdot p - (1-p) \cdot \varepsilon = \varepsilon(2p-1) = \varepsilon \frac{e^\varepsilon - 1}{e^\varepsilon + 1} = \varepsilon \tanh(\varepsilon/2) \leq \frac{\varepsilon^2}{2}$$

Leaky RR

let U, V in $\{0, 1, \text{"I am } U", \text{"I am } V"\}$.

	0	1	"I am U"	"I am V"
U	$(1-\delta)p$	$(1-\delta)(1-p)$	δ	0
V	$(1-\delta)(1-p)$	$(1-\delta)p$	0	δ

Simulation lemma

let P, Q distributions on finite output space. Suppose for every E :

$$P(E) \leq e^\epsilon Q(E) + \delta \quad \text{and} \quad Q(E) \leq e^\epsilon P(E) + \delta$$

Then $\exists$ randomized map F : $F(U) \sim P$ and $F(V) \sim Q$.

Fix $x \sim x'$. Fix j and history $h = y_{<j}$

$P_{j,h}$ is the distribution $A_j(x, h)$

$Q_{j,h}$ " " " " $A_j(x', h)$

By simulation lemma:

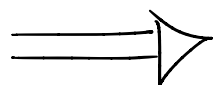
$\exists$ randomized map $F_{j,h}$: $F_{j,h}(U) \sim A_j(x, h)$

$F_{j,h}(V) \sim A_j(x', h)$

Define a randomized map F^* on $z_{1:k} \in \{0, 1, \text{"I am U"}, \text{"I am V"}\}^k$.

$F^*(z_{1:k}) :=$ for $j=1, \dots, k$
| $y_j \sim F_{j, y_{<j}}(z_j)$

Proof by induction



return $y_{1:k}$.

If $(U_1, U_2, \dots, U_k)$ are iid copies of U } then $F^*(U_1, \dots, U_k) \sim A(x)$
 $(V_1, \dots, V_k)$ " " " " V } $F^*(V_1, \dots, V_k) \sim A(x')$

Back to bounding $L_{X \rightarrow X'}(Y_{1:k})$

Remains to show that:

$$L_{U^{\otimes k} \rightarrow V^{\otimes k}}(z) = \log \frac{\prod_{j=1}^k P_U(z_j)}{\prod_{j=1}^k P_V(z_j)} \quad \text{where } P_U(z_j) = \Pr[U = z_j] \\ \text{, } P_V(z_j) = \Pr[V = z_j]$$

$\nwarrow$
 $(z_1, \dots, z_k)$

Since $A(x), A(x')$ are post-processings of $U^{\otimes k}, V^{\otimes k}$, respectively.

let $\text{BAD} = \{ \exists j : z_j = \text{"I am U"} \}$ for $z \sim U^{\otimes k}$

$$\Pr[\text{BAD}] = 1 - (1-\delta)^k \leq k\delta$$

Condition on BAD^c :

$$\text{let } X_j := \log \frac{P_U(z_j)}{P_V(z_j)} = \begin{cases} +\varepsilon & , z_j = 0 \\ -\varepsilon & , z_j = 1 \end{cases} \quad \text{iid random variables}$$

$$L_{U^{\otimes k} \rightarrow V^{\otimes k}}(z) = \sum_{j=1}^k X_j$$

$\mathbb{E}[X_j] = \varepsilon^2/2$, $X_j \in [-\varepsilon, \varepsilon]$ By Hoeffding's.

$$\Pr \left[\sum_{j=1}^k X_j - k\varepsilon^2/2 > t \mid \text{BAD}^c \right] \leq e^{-\frac{t^2}{2\varepsilon^2 k}}$$

set $t = \varepsilon \sqrt{2k \log(1/\delta')}$

$$\Pr \left[L_{U^{\otimes k} \rightarrow V^{\otimes k}}(z) > \underbrace{k\varepsilon^2/2 + \varepsilon \sqrt{2k \log(1/\delta')}}_{\tilde{\varepsilon}} \mid \text{BAD}^c \right] \leq \delta'$$

$$\Rightarrow \Pr \left[L_{U^{\otimes k} \rightarrow V^{\otimes k}}(z) > \tilde{\varepsilon} \right] \leq \delta' + \Pr[\text{BAD}] \leq \delta' + k\delta = \tilde{\delta}.$$

▀

Remarks

- The presentation followed the BU/NEU course. Result by Kairouz, Oh, Viswanath: "The composition theorem of DP." (2015).
- First paper on advanced composition by Dwork, Rothblum, Vadhan: "Boosting and DP." (2010)
- Weaker result (with Azuma!): read Ch. 3.5 Dwork-Roth book
- Notes on simulation lemma: in Murtagh, Vadhan. "The Complexity of Computing the Optimal Composition of DP." (2016).