Differentially Private Learning Needs Hidden State (Or Much Faster Convergence)

Jiayuan Ye (jiayuan@comp.nus.edu.sg), Reza Shokri (reza@comp.nus.edu.sg)
Department of Computer Science, National University of Singapore

Differential Privacy

- Differential Privacy: the distribution of algorithm $\mathscr{A}$'s outputs, on any neighboring inputs, are **indistinguishable**.
- (α, ϵ)-Rényi DP [29]: for any neighboring datasets D, D'

$$R_\alpha(\mathscr{A}(D)\|\mathscr{A}(D')) \leq \epsilon$$

Rényi divergence: $R_\alpha(P\|Q) = \frac{1}{\alpha-1}\log \mathbb{E}_{\theta\sim Q}\left[\left(\frac{P(\theta)}{Q(\theta)}\right)^\alpha\right]$

Standard DP Composition for noisy SGD

- $\theta_0 \leftarrow$ initialization
- Dataset $D = (x_1, \cdots, x_n)$
- For $k = 1, \cdots, K$ do
  - Sample a minibatch B_k
  - $\theta_{k+1} = \text{Update}\left(\theta_k, B_k\right)$ + Noise
- Output θ_K and $\theta_{K-1}, \cdots, \theta_1$

DP Composition Analysis

(α, ϵ) - Rényi DP

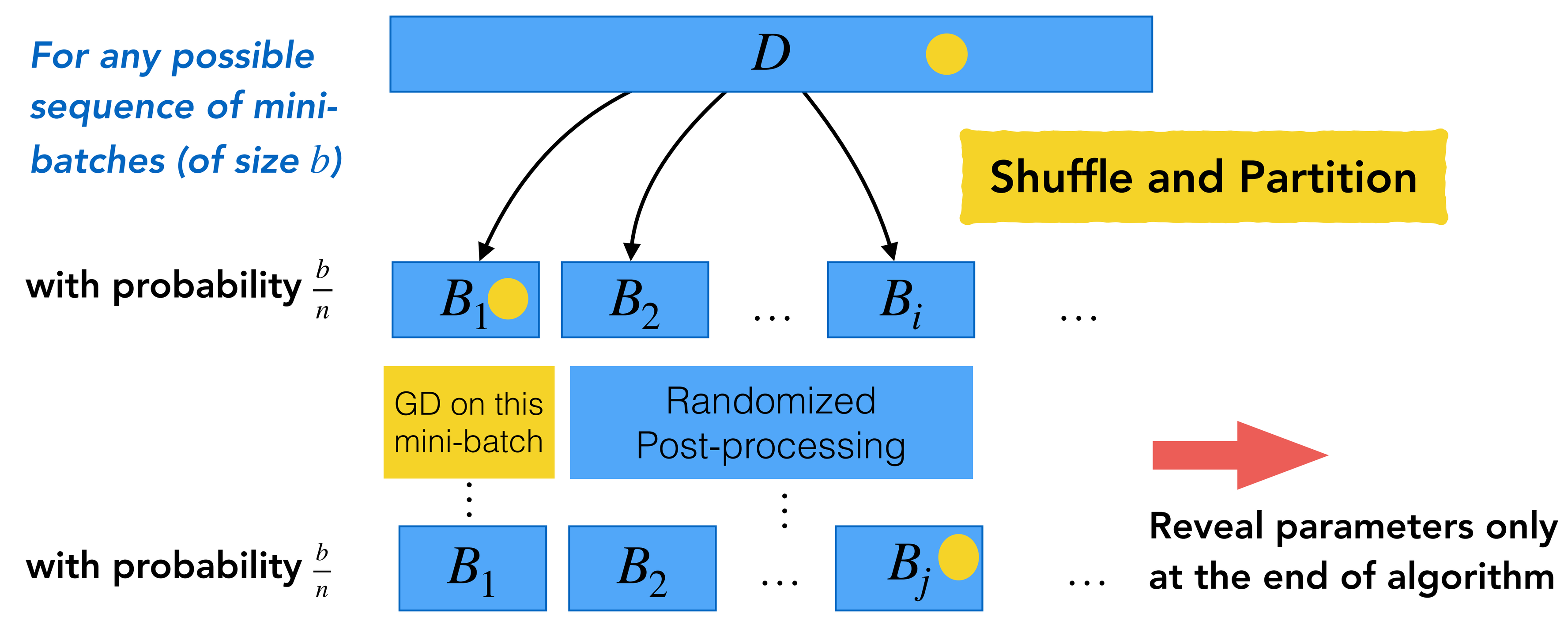
$(\alpha, \epsilon \cdot K)$ - Rényi DP

Has a Complicated Distribution → Idea: $R_\alpha(\theta_K, \cdots, \theta_1 \| \theta'_K, \cdots, \theta_1) \geq R_\alpha\left(\theta_K \| \theta'_K\right)$ by definition

Quantitatively not ideal if the number of iterations K is large

Problem: could we directly prove a (better) DP bound for noisy SGD under the hidden-state assumption (i.e., analyze $R(\theta_K \| \theta'_K)$ while assuming hidden $\theta_{K-1}, \cdots, \theta_1$)?

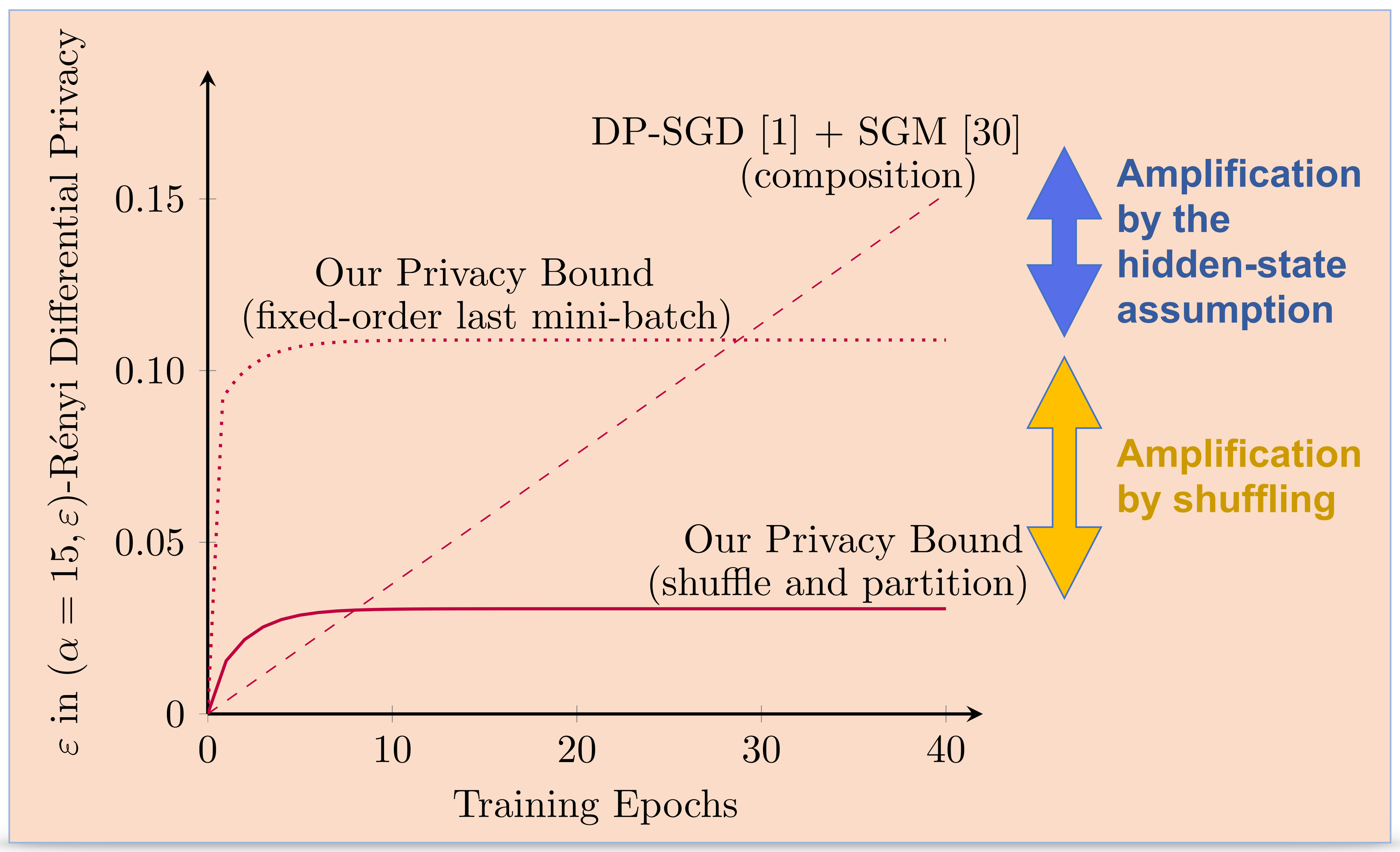
A Better Bound: DP Amplification under Hidden-state Subsampling of noisy SGD



Main Theorem: For λ-strongly convex, β-smooth loss functions with ℓ_2-gradient sensitivity S_g, running Noisy SGD on $\frac{n}{b} \geq 2$ shuffled-once mini-batch partitions with $K \geq 1$ epochs and step-size $\eta < \frac{2}{\lambda+\beta}$ satisfies (α, ε)-Rényi DP with

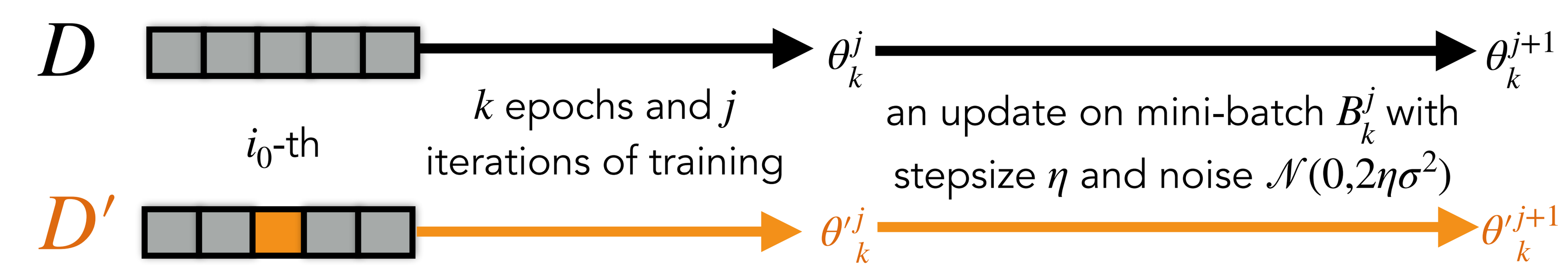
$$\varepsilon \leq \varepsilon_0^{\lfloor \frac{n}{2b} \rfloor}(\alpha) \cdot \frac{1-(1-\eta\lambda)^{2\cdot(K-1)\cdot(n/b-\lfloor \frac{n}{2b} \rfloor)}}{1-(1-\eta\lambda)^{2\cdot(n/b-\lfloor \frac{n}{2b} \rfloor)}} + \frac{1}{\alpha-1}\cdot \log\left(\underset{0\leq j_0<n/b}{Avg} e^{(\alpha-1)\varepsilon_0^{n/b-j_0}(\alpha)}\right)$$

where $\varepsilon_0^j(\alpha) = \frac{\alpha\eta S_g^2}{4\sigma^2 b^2}\cdot(1-\eta\lambda)^{2\cdot(j-1)} \cdot \frac{1}{\sum_{s=0}^{j-1}(1-\eta\lambda)^{2s}}$ for any $j = 1, \cdots, \frac{n}{b}$.



How does hiding intermediate models amplify differential privacy?

*The privacy loss for a mini-batch update is **amplified** if it only accesses every sensitive record with a small probability (due to sub-sampling)*



Theorem: If distributions of θ_k^j and θ'^j_k satisfy log-Sobolev inequality with constant c, and if each mini-batch GD mapping is L-Lipschitz, then

$$\frac{R_\alpha(\theta_k^{j+1}\|\theta'^{j+1}_k)}{\alpha} \leq \begin{cases} \frac{R_{\alpha'}(\theta_k^j\|\theta'^j_k)}{\alpha'}\cdot\left(1+\frac{c\cdot 2\eta\sigma^2}{L^2}\right)^{-1} & \text{if } i_0 \notin B_k^j \\ \frac{R_\alpha(\theta_k^j\|\theta'^j_k)}{\alpha} + \frac{\eta S_g^2}{4\sigma^2 b^2} & \text{if } i_0 \in B_k^j \end{cases} \text{ with } \alpha' = \frac{\alpha-1}{1+\frac{c\cdot 2\eta\sigma^2}{L^2}} + 1.$$

How does stochastic ordering of mini-batches further amplify privacy?

*The privacy loss for a mini-batch update is **amplified** if it only accesses every sensitive record with a small probability (due to sub-sampling)*

Key difficulty: the distribution of the final output is a mixture distribution with *a large number of* mixture components

Our Technique: recursively study divergence between mixtures

$$e^{(\alpha-1)\cdot R_\alpha\left(\sum_{j=1}^m p_j\mu_j \| \sum_{j=1}^m p_j\nu_j\right)} \leq \sum_{j=1}^m p_j \cdot e^{(\alpha-1)\cdot R_\alpha(\mu_j\|\nu_j)}$$

Main Takeaway

- We prove a novel converging last-iterate privacy bound for noisy SGD on strongly convex smooth loss functions.
- Our bound substantially improves over prior privacy bounds, via novel bounds for the additional DP amplification in noisy SGD
- Our results show that to obtain tighter privacy bound, *DP learning algorithms needs to be evaluated by a last-iterate privacy bound, unless it has a very fast convergence.*

[1] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In: ACM CCS 2016.

[29] Ilya Mironov. Rényi differential privacy. In: 2017 IEEE 30th Computer Security Foundations Symposium (CSF), p. 263-275. IEEE, 2017.

[30] Ilya Mironov, Kunal Talwar, and Li Zhang. Rényi differential privacy of the sampled Gaussian mechanism. In: arXiv preprint arXiv: 1908.10530.