Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS 2021)

Differential Privacy Dynamics of Langevin Diffusion and Noisy Gradient Descent

Rishav Chourasia*, Jiayuan Ye*, Reza Shokri

Data Privacy and Trustworthy ML Research Lab
National University of Singapore (NUS)

Privacy Risks of ML Algorithms

Privacy Risk: output model leaks information about the **individual members** of its training dataset

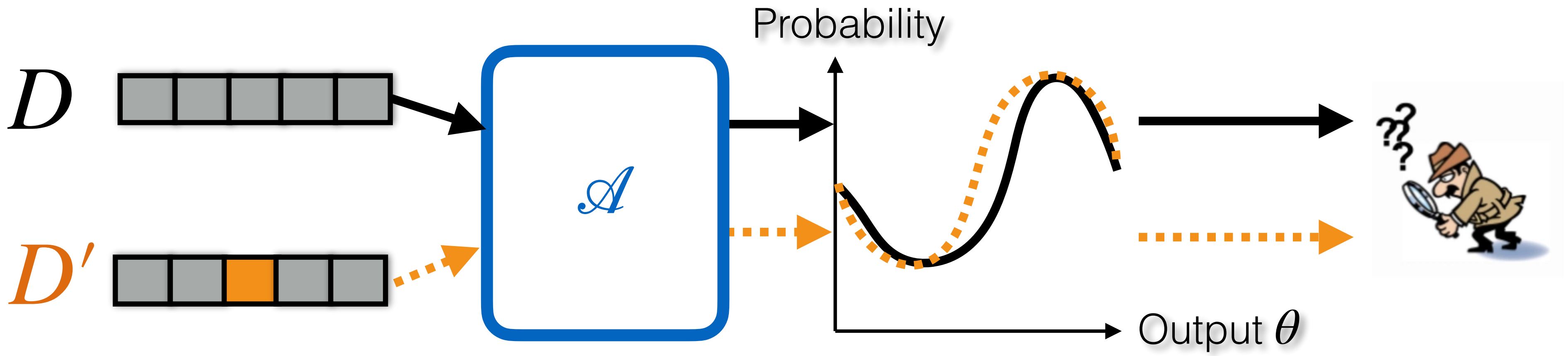
- Membership inference attacks
  - Shokri, Stronati, Song, Shmatikov (2017)
- Reconstruction attacks
  - Carlini, Tramèr, et al. (2021)

Differential Privacy

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

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

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



[Mironov] Rényi differential privacy. CSF 2017

How to Train Privacy-preserving Model

- $\theta_0 \leftarrow$ initialization
- Dataset $D = (x_1, \cdots, x_n)$
- For $k = 1, \cdots, K$ do
  - $\theta_{k+1} =$ Update $\left(\theta_k, D\right)$ **+ Noise**
- Output θ_K

Has a Complicated Distribution

Problem: how to bound the Rényi privacy loss $R_\alpha(\theta_K \| \theta'_K)$

[Mironov] Rényi differential privacy. CSF 2017

How to Train Privacy-preserving Model

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

DP Composition Analysis

(α, ϵ) - Rényi DP

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

$\geq$

Problem: how to bound the Rényi privacy loss $R_\alpha(\theta_K \| \theta'_K)$

[Mironov] Rényi differential privacy. CSF 2017

How to Compute a Better Bound

- A new privacy analysis for the Noisy Gradient Descent on a certain class of loss functions
  - analyzes the privacy loss for revealing the final model θ_K
  - assumes hidden intermediate models $\theta_1, \cdots, \theta_{K-1}$

Input: Dataset $\mathcal{D} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$, loss function ℓ, learning rate η, noise variance σ^2, initial parameter vector θ_0.

1: **for** $k = 0, 1, \cdots, K-1$ **do**

2: $\quad g(\theta_k; \mathcal{D}) = \sum_{i=1}^{n} \nabla \ell(\theta_k; \mathbf{x_i})$

3: $\quad \theta_{k+1} = \Pi_{\mathcal{C}} \left(\theta_k - \frac{\eta}{n} g(\theta_k; D) + \sqrt{2\eta\sigma^2} \mathcal{N}(0, \mathbb{I}_d) \right)$

4: Output θ_K

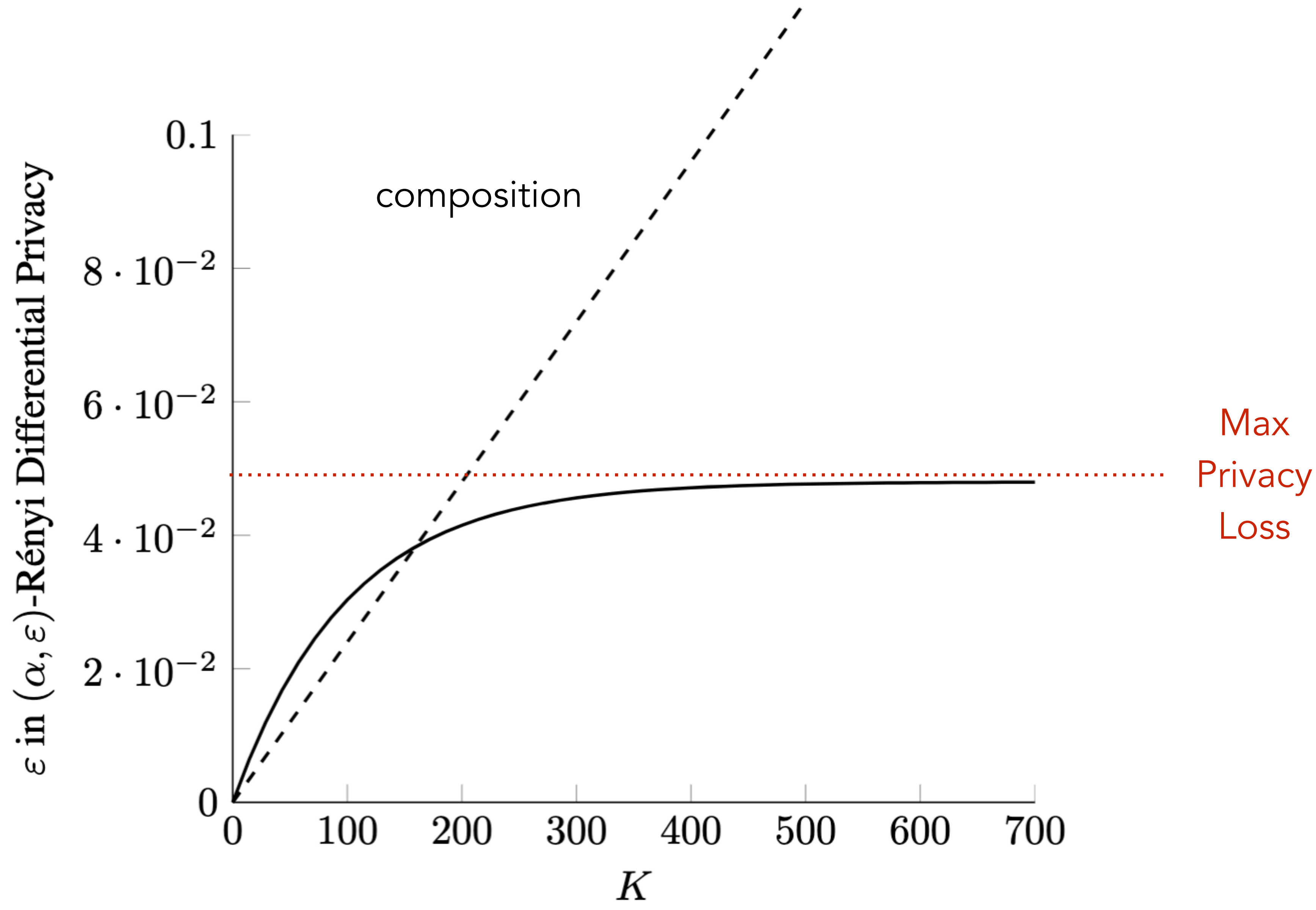
Privacy Dynamics Bound

- **Main Theorem:** Noisy GD on λ-strongly convex β-smooth loss functions with gradient sensitivity $S_g = \max_{D,D'} \|g(\theta; D) - g(\theta; D')\|_2$ step-size $\eta \leq 1/\beta$ and K iterations satisfies (α, ϵ)-Rényi DP

$$\epsilon = \frac{\alpha S_g^2}{\lambda \sigma^2 n^2} \cdot \left(1 - e^{-\lambda \eta K/2}\right)$$

Max Privacy Loss

Privacy Loss Convergence Rate



Parameters: $\alpha = 30$, $\sigma = 0.02$, $S_g = 4$, $\eta = 0.02$, $\lambda = 1$, Size of dataset: $n = 5000$

Our Privacy Analysis is Tight

- **<u>Exact Privacy Loss Lower Bound</u>**
  compute exact privacy loss for noisy GD on the squared norm loss function $\ell(\theta; x) = \|\theta - x\|^2/2$

$$\epsilon \geq \frac{\alpha S_g^2}{4\sigma^2 n^2} \cdot \left(1 - e^{-\eta K}\right)$$

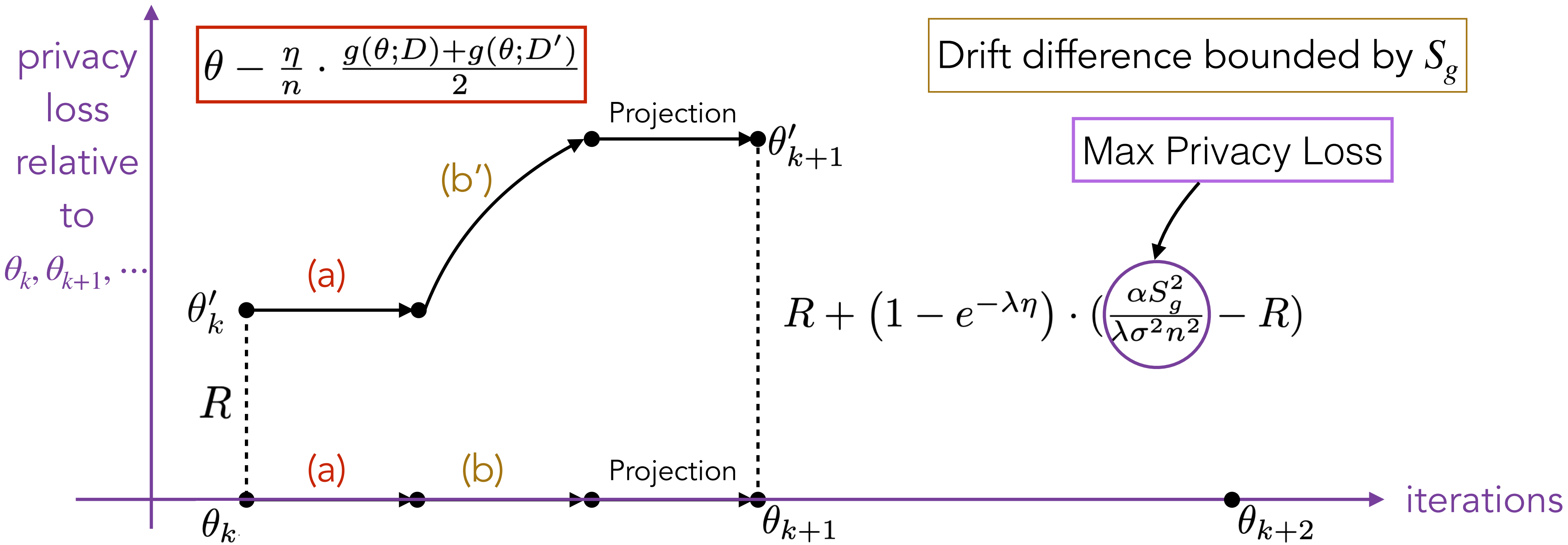
- **<u>Privacy Dynamics Bound</u>**

$$\epsilon = \frac{\alpha S_g^2}{\lambda\sigma^2 n^2} \left(1 - e^{-\lambda\eta K/2}\right)$$

- **<u>Tightness:</u>** the upper bound matches the lower bound up to a small constant of 4

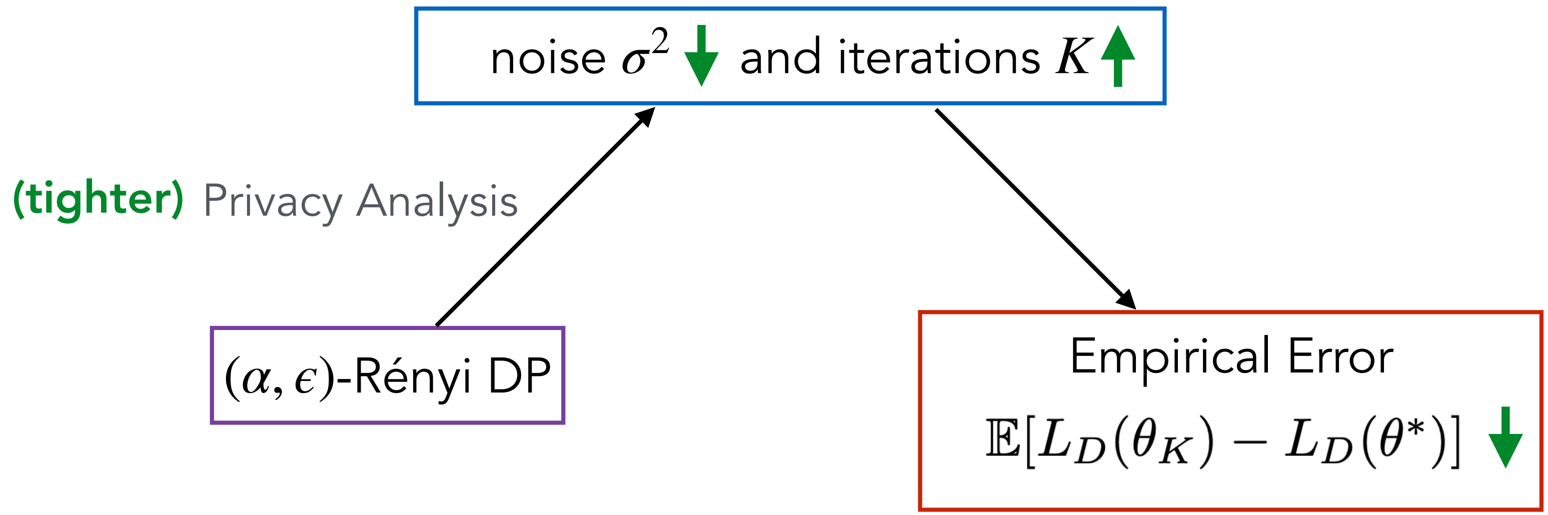
How to Prove Privacy Dynamics

- One Update: $\theta_{k+1} = \Pi_{\mathcal{C}}\left(\theta_k - \frac{\eta}{n} g(\theta_k; D) + \sqrt{2\eta\sigma^2}\mathcal{N}(0, \mathbb{I}_d)\right)$
  - (b) Langevin diffusion with drift $-\frac{1}{n} \cdot \frac{g(\theta_k;D) - g(\theta_k;D')}{2}$
  - (b') Langevin diffusion with drift $-\frac{1}{n} \cdot \frac{g(\theta'_k;D') - g(\theta'_k;D)}{2}$



Utility Analysis

- How does the added randomness required for achieving privacy by a privacy analysis affect the error of the algorithm's output?



Utility Analysis

- Privacy dynamics analysis facilitates a better privacy-utility tradeoff than the DP composition analysis for strongly convex smooth loss functions.

$$\mathbb{E}[L_D(\theta_{K^*}) - L_D(\theta^*)] \leq \frac{\alpha}{\epsilon} \cdot \frac{\beta d L^2}{\lambda^2 n^2}$$

$poly(n)$ smaller runtime

$poly \log n$ smaller error

Summary

- We need better estimates of the privacy loss for differentially-private machine learning algorithms
  - How much does a trained model leak about its training data? Assuming that intermediate steps of the training algorithm are private and not visible to adversary.
- We present a new tight converging privacy dynamics theorem for noisy gradient descent algorithms on strongly convex smooth loss functions
- Open problem: Privacy dynamics under relaxed conditions