

Privacy Analysis of Iterative Algorithms via f -divergences

Shahab Asoodeh

shahab@seas.harvard.edu

Based on

- S. A., M. Diaz, and F. P. Calmon, [Privacy Analysis of Online Learning Algorithms via Contraction Coefficients](#), arXiv:2012.11035
- S. A., J. Liao, F. P. Calmon, O. Kosut, L. Sankar, [Three Variants of Differential Privacy: Lossless Conversion and Applications](#), arXiv:2008.06529



Joint work with: Flavio Calmon, Mario Diaz, Jiachun Liao, Oliver Kosut,
and Lalitha Sankar



HARVARD
John A. Paulson
School of Engineering
and Applied Sciences

Main goals

Optimal Conversion from Rényi DP to Approximate DP

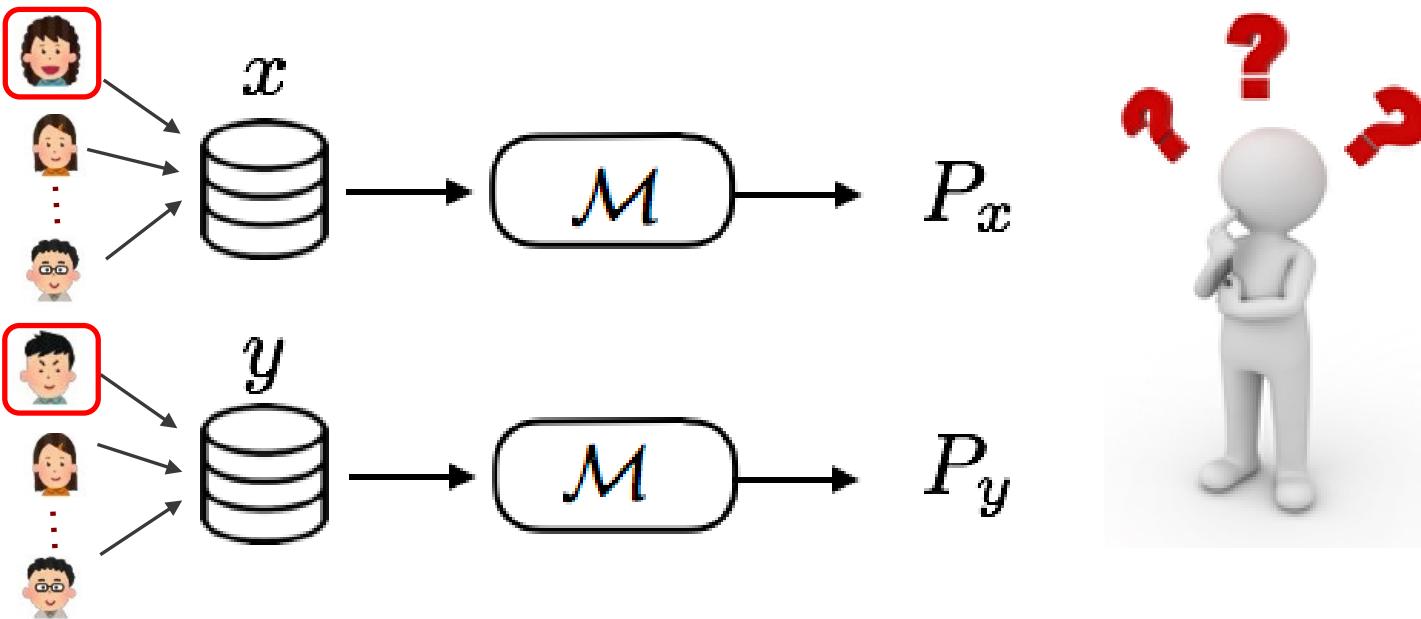
If \mathcal{M} is (α, ζ) -RDP, then what are the smallest ε and δ s.t. it is (ε, δ) -DP?

DP Guarantees of Iterative Algorithms

DP guarantee of releasing W_T the output of $W_{t+1} = \Psi_t(W_t) + Z_{t+1}$ after T iterations?

[assuming W_1, \dots, W_{T-1} are kept hidden]

Differential Privacy



\mathcal{M} is called (ε, δ) -DP if

$$\sup_{x \sim y} \sup_A [P_x(A) - e^\varepsilon P_y(A)] \leq \delta$$

\mathcal{M} is called (α, ζ) -RDP for $\alpha > 1$ if

$$\sup_{x \sim y} D_\alpha(P_x \| P_y) \leq \zeta$$

Why conversion?

DP



Operationality: Hypothesis testing

- [Wasserman and Zhou'10], [Kairouz et al.'15]



Composability

- [Kairouz et al.'15], [Dwork et al.'10]
[Murtagh and Vadhan'15]

RDP



Composability

- [Abadi et al.'16], [Mironov'17],
[Bun et al.'18], [Bun and Steinke'16]



Operationality

- [Balle et al.'19]

Moments accountant*

[Abadi et al, 2016]

For an iterative alg \mathcal{M} :

Find $\zeta_i(\alpha)$ s.t. i th iteration is \Rightarrow After n iterations \mathcal{M} is \Rightarrow It is (ε, δ) -DP for any ε and
 $(\alpha, \zeta_i(\alpha))$ -RDP $\forall \alpha > 1$ $(\alpha, \zeta(\alpha))$ -RDP where
$$\zeta(\alpha) = \sum_{i=1}^n \zeta_i(\alpha)$$

 $\delta = e^{-(\alpha-1)(\varepsilon-\zeta(\alpha))}$

or for any δ and

$$\varepsilon = \inf_{\alpha > 1} \zeta(\alpha) - \frac{\log \delta}{\alpha - 1}$$

To develop the optimal conversion rule, we first fix α :

$$\zeta(\alpha) \longleftrightarrow \zeta$$

$$\delta_\alpha^\varepsilon(\zeta) \triangleq \inf \left\{ \delta \in [0, 1] : \forall (\alpha, \zeta)\text{-RDP } \mathcal{M} \text{ are } (\varepsilon, \delta)\text{-DP} \right\}$$

$$\varepsilon_\alpha^\delta(\zeta) \triangleq \inf \left\{ \varepsilon \geq 0 : \forall (\alpha, \zeta)\text{-RDP } \mathcal{M} \text{ are } (\varepsilon, \delta)\text{-DP} \right\}$$

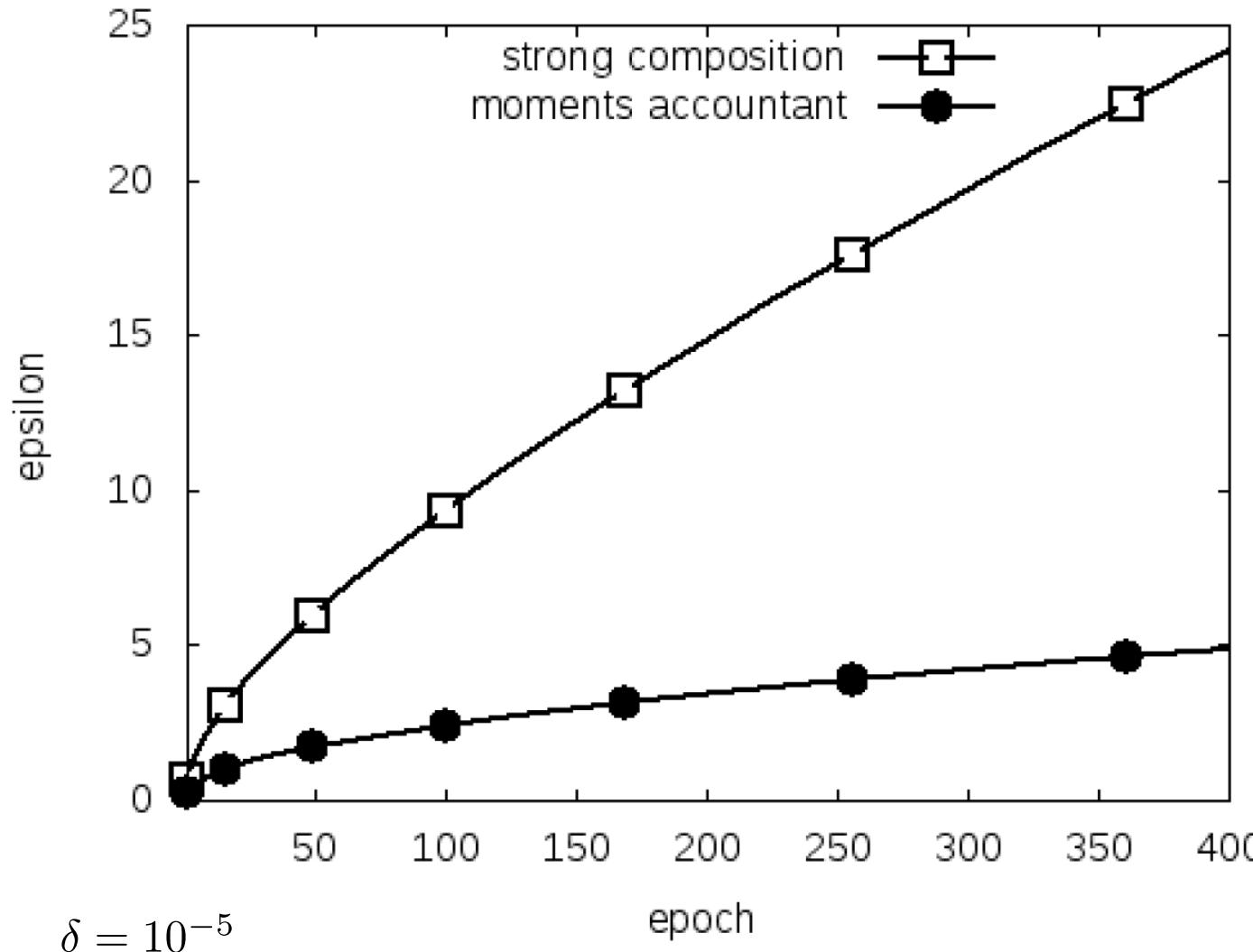
$$\zeta_\alpha^\varepsilon(\delta) \triangleq \sup \{ \zeta \geq 0 : \forall (\alpha, \zeta)\text{-RDP } \mathcal{M} \text{ are } (\varepsilon, \delta)\text{-DP} \}$$



RDP \longrightarrow DP

DP \longrightarrow RDP

Moments accountant for noisySGD



How much more improvement if we “optimize” the RDP to DP conversion?

A better conversion rule gives 100 more iterations!

Differential privacy as f -divergence

Given convex function f satisfying $f(1) = 0$: $D_f(P\|Q) \triangleq \mathbb{E}_Q \left[f\left(\frac{dP}{dQ}\right) \right]$

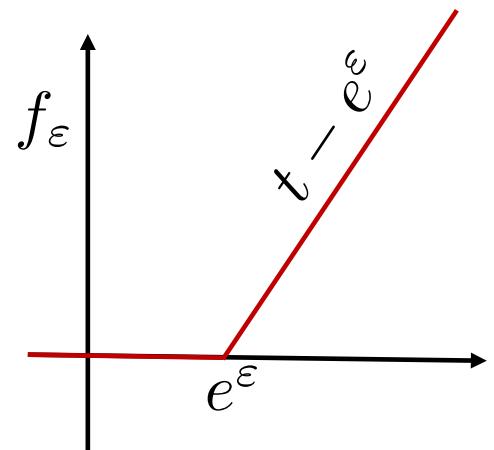
- $f_\alpha(t) = \frac{t^\alpha - 1}{\alpha - 1}$ and let $\chi^\alpha(P\|Q) \triangleq D_{f_\alpha}(P\|Q)$

$$\zeta$$

$$\mathcal{M} \text{ is } (\alpha, \zeta)\text{-RDP} \iff \sup_{x \sim y} \chi^\alpha(P_x\|P_y) \leq \frac{1}{\alpha - 1} [e^{(\alpha-1)\zeta} - 1]$$

- $f_\varepsilon(t) \triangleq \max(t - e^\varepsilon, 0)$ and let $\mathsf{E}_\varepsilon(P\|Q) \triangleq D_{f_\varepsilon}(P\|Q)$

$$\mathcal{M} \text{ is } (\varepsilon, \delta)\text{-DP} \iff \sup_{x \sim y} \mathsf{E}_\varepsilon(P_x\|P_y) \leq \delta$$



Problem formulation

Given $\alpha > 1$, let

$$\mathbb{M}_\alpha(\zeta) = \{\text{all } (\alpha, \zeta)\text{-RDP mechanisms}\}$$

Given $\varepsilon \geq 0$, we wish to characterize

$$\delta_\alpha^\varepsilon(\zeta) := \inf \{\delta \in [0, 1] : \forall \mathcal{M} \in \mathbb{M}_\alpha(\zeta) \text{ is } (\varepsilon, \delta)\text{-DP}\}$$

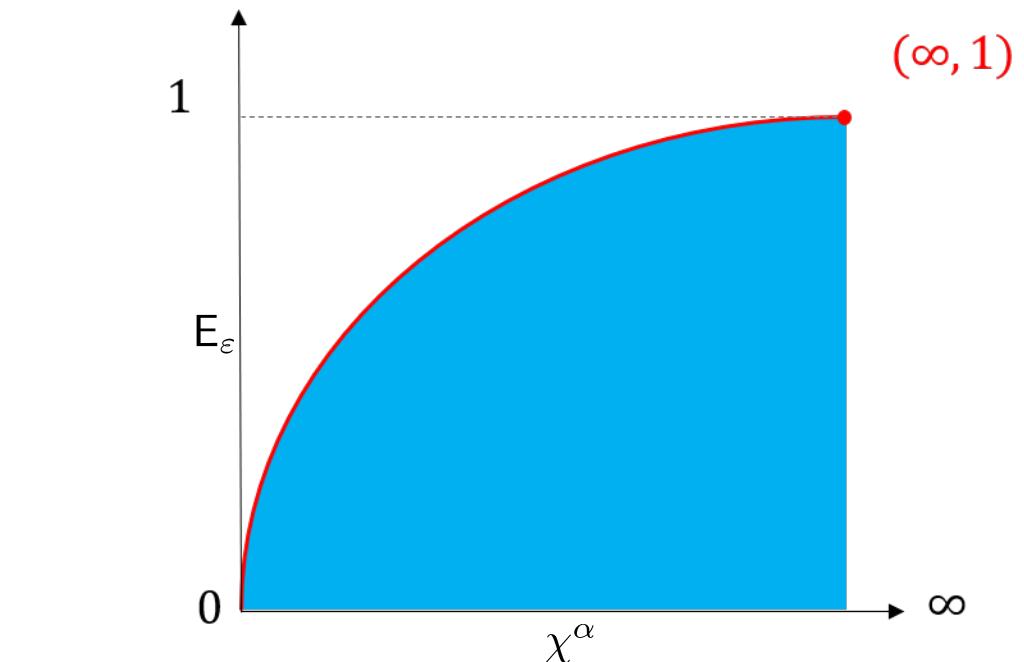
$$= \sup_{\mathcal{M} \in \mathbb{M}_\alpha(\zeta)} \sup_{x \sim y} \mathsf{E}_\varepsilon(P_x \| P_y)$$

$$= \sup_{\mathcal{M}} \sup_{x \sim y} \mathsf{E}_\varepsilon(P_x \| P_y)$$

$$\text{s.t. } \chi^\alpha(P_x \| P_y) \leq \tilde{\zeta}, \forall x \sim y$$

$$= \sup_{P, Q} \mathsf{E}_\varepsilon(P \| Q)$$

$$\text{s.t. } \chi^\alpha(P \| Q) \leq \tilde{\zeta}$$



$$\mathcal{R}_\alpha := \{(\chi^\alpha(P \| Q), \mathsf{E}_\varepsilon(P \| Q)) : \forall \text{dis. } P, Q\}$$

Joint range of f-divergences

- Characterize $R_\alpha = \{(\chi^\alpha(P\|Q), E_\varepsilon(P\|Q)), \forall P, Q\}$

- Derive the upper boundary of R_α

- $\tilde{\zeta} \rightarrow \zeta$

Theorem [Harremöes, Vajda'11]. Given two convex function f and g with $f(1) = g(1) = 0$, we have

$$\left\{ (D_f(P\|Q), D_g(P\|Q)) : \forall P, Q \right\} = \text{conv}(\mathcal{B}),$$

where

$$\mathcal{B} := \left\{ (D_f(\text{Ber}(p)\|\text{Ber}(q)), D_g(\text{Ber}(p)\|\text{Ber}(q))) : p, q \in (0, 1) \right\}.$$

$$\mathcal{R}_\alpha = \text{conv} \left(\left\{ (\chi^\alpha(\text{Ber}(p)\|\text{Ber}(q)), E_\varepsilon(\text{Ber}(p)\|\text{Ber}(q))) : p, q \in (0, 1) \right\} \right).$$

Upper boundary:

$$\begin{aligned} & \sup_{p, q \in (0, 1)} E_\varepsilon(\text{Ber}(p)\|\text{Ber}(q)) & \leftrightarrow & \inf_{p, q \in (0, 1)} \chi^\alpha(\text{Ber}(p)\|\text{Ber}(q)) & = \zeta_\alpha^\varepsilon(\delta) \\ & \text{s.t. } \chi^\alpha(\text{Ber}(p)\|\text{Ber}(q)) \leq \tilde{\zeta} & & \text{s.t. } E_\varepsilon(\text{Ber}(p)\|\text{Ber}(q)) \geq \delta \end{aligned}$$

Main result 1

Theorem. For any $\alpha > 1$, $\varepsilon \geq 0$, we have

$$\zeta_\alpha^\varepsilon(\delta) = \varepsilon + \frac{1}{\alpha - 1} \log M(\alpha, \varepsilon, \delta)$$

where

$$M(\alpha, \varepsilon, \delta) \triangleq \min_{p \in (\delta, 1)} [p^\alpha (p - \delta)^{1-\alpha} + (1 - p)^\alpha (e^\varepsilon - p + \delta)^{1-\alpha}] .$$

$$\zeta_\alpha^\varepsilon(\delta) \iff M(\alpha, \varepsilon, \delta)$$

∞ -dim non-convex
optimization problem

one-dim convex
optimization problem

Main result 2

Theorem. For any $\varepsilon \geq 0$ and $\alpha > 1$, we have

$$\zeta_\alpha^\varepsilon(0) = 0,$$

$$\zeta_\alpha^\varepsilon(\delta) = \varepsilon - \log(1 - \delta), \quad \text{if } \alpha\delta \geq 1,$$

Main result 2

Theorem. For any $\varepsilon \geq 0$ and $\alpha > 1$, we have

$$\zeta_\alpha^\varepsilon(0) = 0,$$

$$\zeta_\alpha^\varepsilon(\delta) = \varepsilon - \log(1 - \delta), \quad \text{if } \alpha\delta \geq 1,$$

$$\zeta_\alpha^\varepsilon(\delta) \geq \max\{g(\alpha, \varepsilon, \delta), f(\alpha, \varepsilon, \delta)\}, \quad \text{if } 0 < \alpha\delta < 1,$$

where

$$g(\alpha, \varepsilon, \delta) := \varepsilon - \frac{1}{\alpha - 1} \log \frac{h_\alpha}{\delta},$$

with $h_\alpha := \frac{1}{\alpha} \left(1 - \frac{1}{\alpha}\right)^{\alpha-1}$ and

$$f(\alpha, \varepsilon, \delta) := \varepsilon + \frac{1}{\alpha - 1} \log \left((e^\varepsilon - \alpha\delta) \left(\frac{\delta - 1}{\delta - e^\varepsilon} \right)^\alpha + \alpha\delta \right).$$

By linearizing f , we can obtain upper bound for $\delta_\alpha^\varepsilon(\zeta)$ or $\varepsilon_\alpha^\delta(\zeta)$:

Otherwise: $\varepsilon_\alpha^\delta(\zeta) \leq \frac{1}{\alpha-1} \min \left\{ (\alpha - 1)\zeta - \log \frac{\delta}{h_\alpha}, \log \left(\frac{e^{(\alpha-1)\zeta} - 1}{\alpha\delta} + 1 \right) \right\}$

Related work

$$\varepsilon_\alpha^\delta(\zeta) \begin{cases} = \zeta + \log(1 - \delta) & \text{if } \alpha\delta \geq 1 \\ \leq \frac{1}{\alpha-1} \min \left\{ (\alpha - 1)\zeta - \log \frac{\delta}{h_\alpha}, \log \left(\frac{e^{(\alpha-1)\zeta} - 1}{\alpha\delta} + 1 \right) \right\} & \text{otherwise} \end{cases}$$

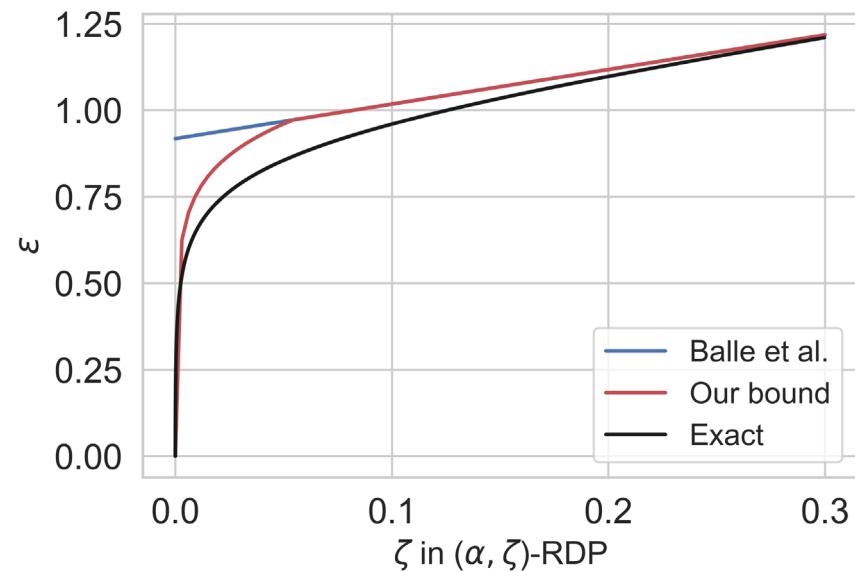
[Balle et al.'20] and [Canonne et al.'20] recently showed

$$(\alpha, \zeta)\text{-RDP} \Rightarrow \left(\zeta - \frac{1}{\alpha-1} \log \frac{\delta}{h_\alpha}, \delta \right)\text{-DP}$$

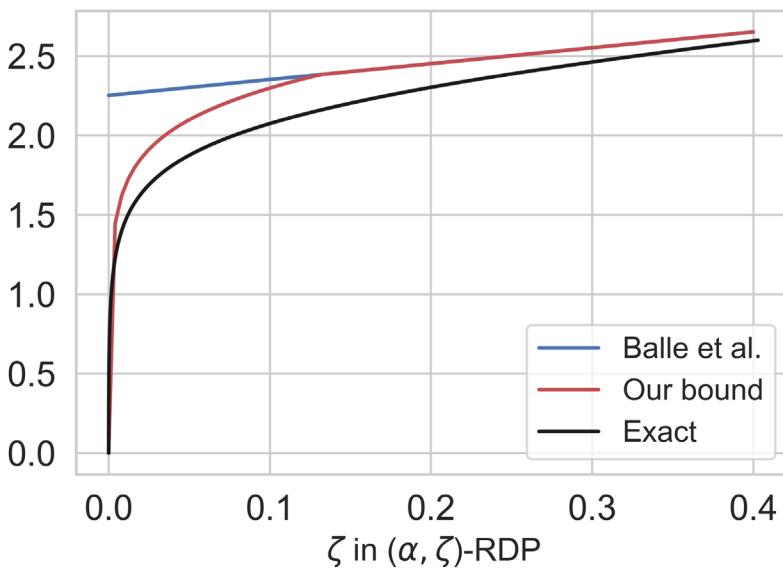
Balle, Barthe, Gaboardi, Hsu, Sato, "Hypothesis testing interpretations and Rényi differential privacy", AISTAT2020

Canonne, Kamath, and Steinke, "The discrete Gaussian for differential privacy", arXiv:2004.00010, 2020

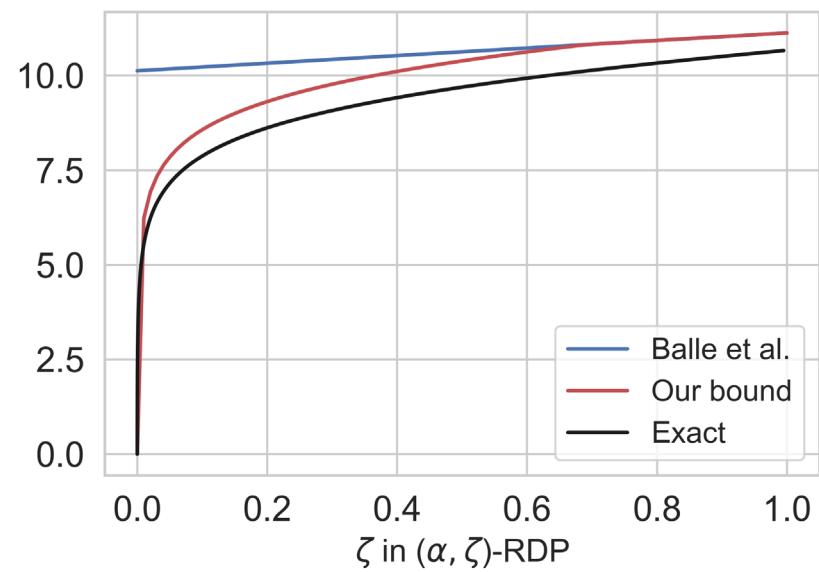
Comparison



$$\alpha = 10, \delta = 10^{-5}$$



$$\alpha = 5, \delta = 10^{-5}$$



$$\alpha = 2, \delta = 10^{-5}$$

$$\varepsilon_\alpha^\delta(\zeta) \begin{cases} = \zeta + \log(1 - \delta) \\ \leq \frac{1}{\alpha-1} \min \left\{ (\alpha - 1)\zeta - \log \frac{\delta}{h_\alpha}, \log \left(\frac{e^{(\alpha-1)\zeta} - 1}{\alpha\delta} + 1 \right) \right\} \end{cases} \quad \begin{array}{l} \text{if } \alpha\delta \geq 1 \\ \text{otherwise} \end{array}$$

Improved moments accountant

Stochastic Gradient Descent

Given dataset: $x = \{x_1, \dots, x_N\}$

- $\theta_0 \leftarrow$ initial randomly
- for $i = 1$ to T do

Take random samples B of size qN from x

$$\theta_i \leftarrow \left(\theta_{i-1} - \eta \left(\frac{1}{qN} \sum_{j \in B} \nabla \ell(\theta_{i-1}, x_j) + Z_i \right) \right)$$

- return θ_T

$\mathcal{N}(0, \sigma^2 I)$

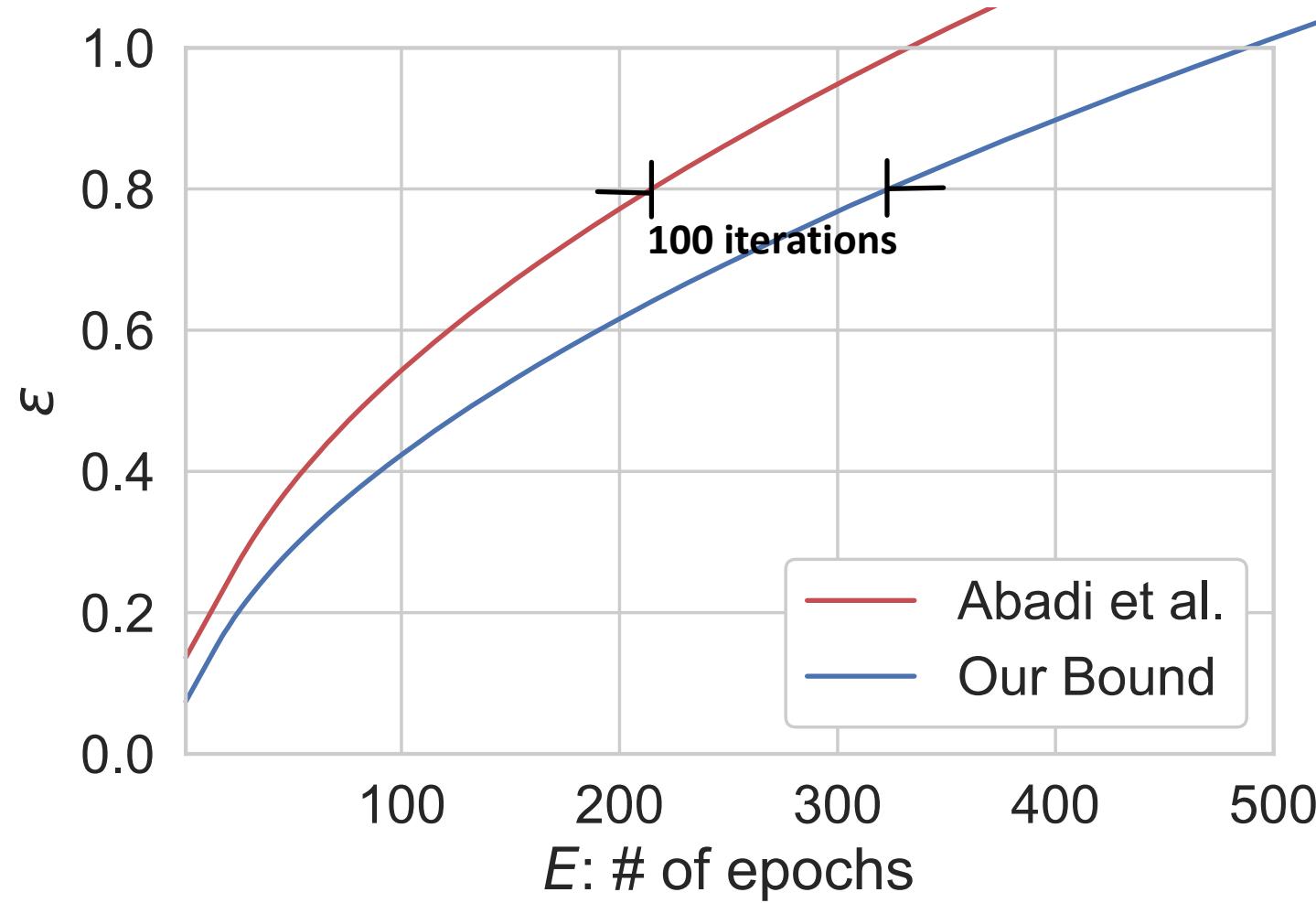
- [Abadi et al.'16]: above alg. is $(\alpha, \alpha\rho T)$ -RDP with $\rho = \frac{q^2}{(1-q)\sigma^2}$
- After T iterations, it is (ε, δ) -DP

[Abadi et al'16] $\varepsilon = \inf_{\alpha > 1} \left[\alpha \rho T - \frac{1}{\alpha - 1} \log \delta \right] = \rho T + \sqrt{4\rho T \log \frac{1}{\delta}}$

[Our bound] $\varepsilon = \inf_{\alpha > 1} \frac{1}{\alpha - 1} \min \left\{ (\alpha - 1)\alpha \rho T - \log \frac{\delta}{h_\alpha}, \log \left(\frac{e^{(\alpha-1)\alpha \rho T} - 1}{\alpha \delta} + 1 \right) \right\}$

Improved moments accountant

$q = 0.04, \sigma = 4$



$$\varepsilon = \rho T + \sqrt{4\rho T \log \frac{1}{\delta}}$$

$$\varepsilon = \inf_{\alpha > 1} \frac{1}{\alpha - 1} \min \left\{ (\alpha - 1)\alpha \rho T - \log \frac{\delta}{h_\alpha}, \log \left(\frac{e^{(\alpha-1)\alpha \rho T} - 1}{\alpha \delta} + 1 \right) \right\}$$

Main Goals



Optimal Conversion from Rényi DP to Approximate DP

If \mathcal{M} is (α, ζ) -RDP, then what are the smallest ε and δ s.t. it is (ε, δ) -DP?

DP Guarantees of Iterative Algorithms

DP guarantee of releasing W_T the output of $W_{t+1} = \Psi_t(W_t) + Z_{t+1}$ after T iterations?

[assuming W_1, \dots, W_{T-1} are kept hidden]

Privacy amplification by iteration*

[Feldman et al., 2018]

Consider the process:

$$W_{t+1} = \Psi_{t+1}(W_t) + Z_{t+1},$$

for some sequence of update functions $\{\Psi_t\}$ and $Z_t \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_d)$.



Theorem [Feldman et al.]. If $\{\Psi_t\}$ are **contractive**, then for every $\alpha > 1$,

$$D_\alpha(\mu_T \| \mu'_T) \leq \frac{\alpha}{2\sigma^2 T} \|w_0 - w'_0\|_2^2.$$

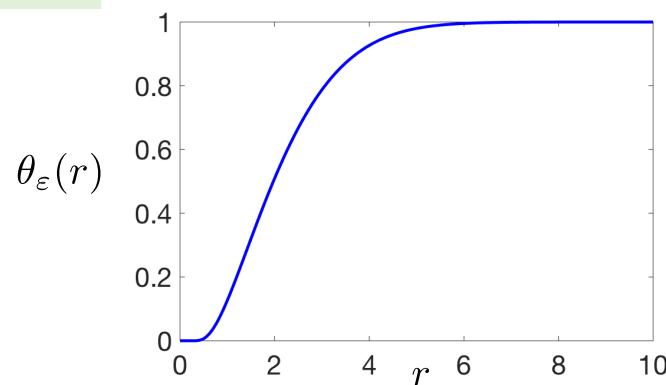
Our result



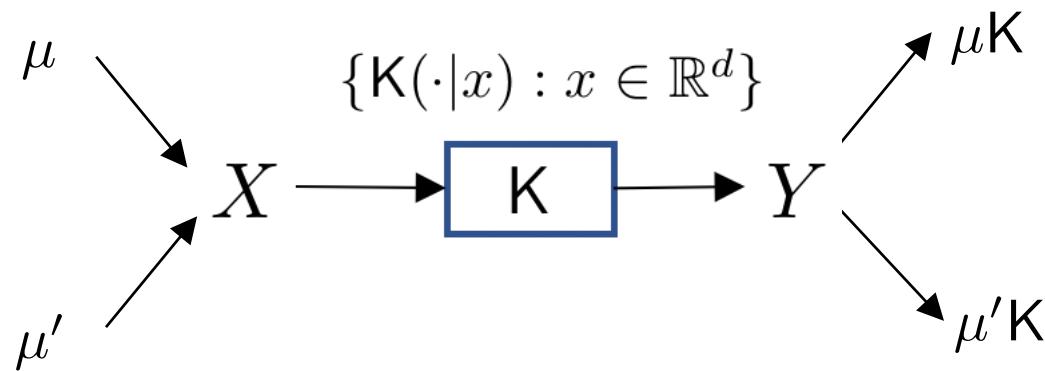
Theorem. Assume $\|\Psi_t(w) - \Psi_t(w')\| \leq D$ for all t and w, w' . Then we have

$$\mathsf{E}_\varepsilon(\mu_T \| \mu'_T) \leq \mathsf{E}_\varepsilon(\mu_1 \| \mu'_1) \left[\theta_\varepsilon \left(\frac{D}{\sigma} \right) \right]^T,$$

where $\theta_\varepsilon(r) \triangleq Q \left(\frac{\varepsilon}{r} - \frac{r}{2} \right) - e^\varepsilon Q \left(\frac{\varepsilon}{r} + \frac{r}{2} \right)$ and $Q(a) \triangleq \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-u^2/2} du$.



SDPI

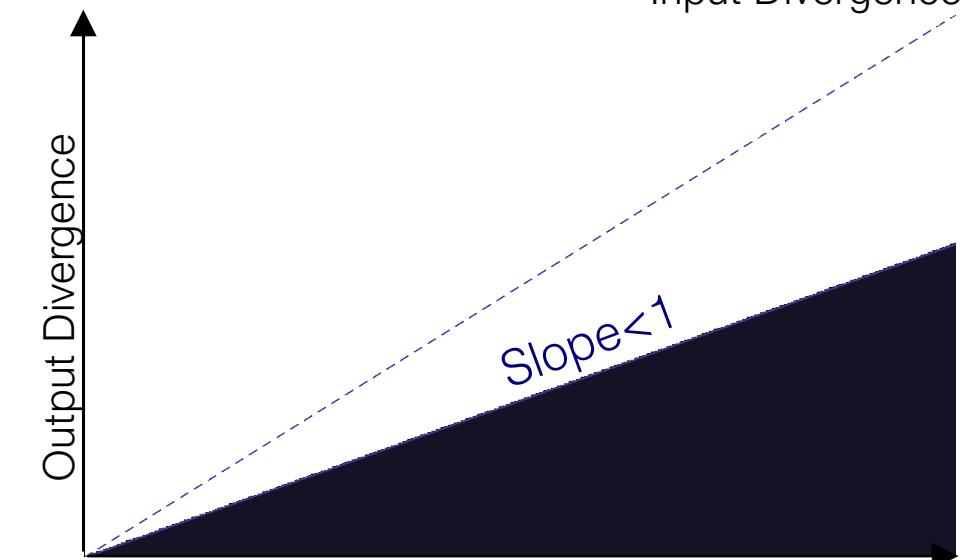
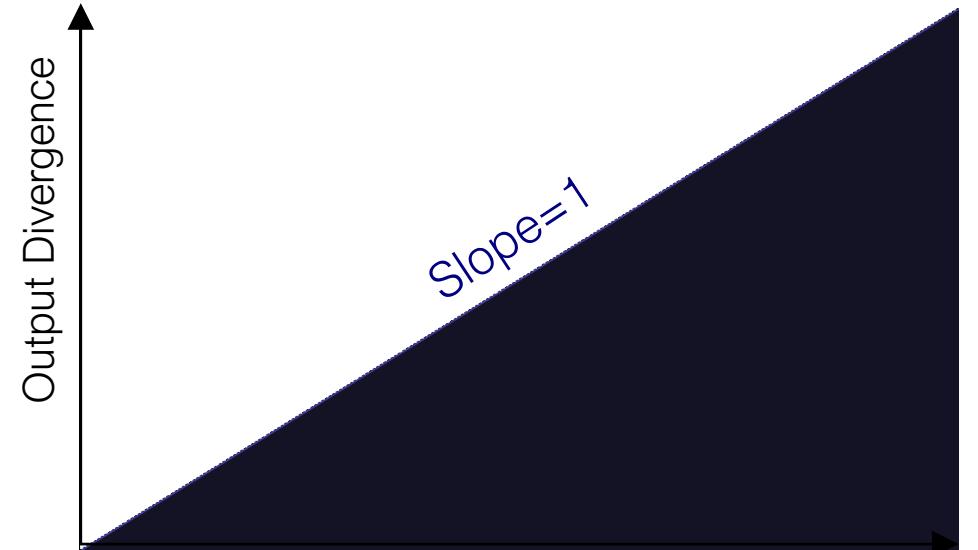


$$\text{DPI: } \frac{D_f(\mu K \| \mu' K)}{D_f(\mu \| \mu')} \leq 1$$

$$\text{strong DPI: } \sup_{\substack{\mu, \mu' \\ D_f(\mu \| \mu') \neq 0}} \frac{D_f(\mu K \| \mu' K)}{D_f(\mu \| \mu')}$$

Contraction coefficient of K under f -divergence

$$D_f(\mu K \| \mu' K) \leq D_f(\mu \| \mu') \eta_f(K)$$



SDPI

Total variation:

$$\eta_{\text{TV}}(K) \triangleq \sup_{\mu \neq \mu'} \frac{\text{TV}(\mu K, \mu' K)}{\text{TV}(\mu, \mu')} = \max_{x, x'} \text{TV}(K(\cdot|x), K(\cdot|x'))$$

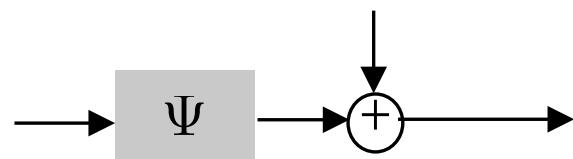
[Dobrushin's Theorem]

[Dobrushin'56]

Theorem. Given a Markov kernel K and $\varepsilon \geq 0$, we have

$$\eta_\varepsilon(K) \triangleq \sup_{\substack{\mu, \mu' \\ E_\varepsilon(\mu \| \mu') \neq 0}} \frac{E_\varepsilon(\mu K \| \mu' K)}{E_\varepsilon(\mu \| \mu')} = \max_{x, x'} E_\varepsilon(K(\cdot|x) \| K(\cdot|x')).$$

$$Z \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_d)$$



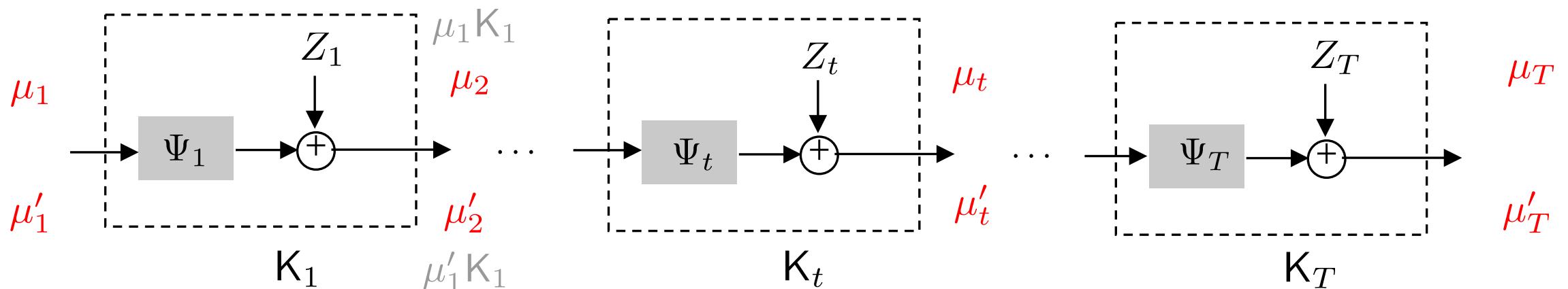
$$\eta_\varepsilon(K_{\text{Gaussian}}) = \theta_\varepsilon\left(\frac{D}{\sigma}\right) \quad \text{if } \|\Psi(x) - \Psi(x')\| \leq D, \forall x, x'$$

$$K_{\text{Gaussian}}$$

$$\theta_\varepsilon(r) \triangleq Q\left(\frac{\varepsilon}{r} - \frac{r}{2}\right) - e^\varepsilon Q\left(\frac{\varepsilon}{r} + \frac{r}{2}\right)$$

Proof: SDPI

$$W_{t+1} = \Psi_{t+1}(W_t) + Z_{t+1}$$



$$\mathsf{E}_\varepsilon(\mu_T \| \mu'_T) \leq \mathsf{E}_\varepsilon(\mu_{T-1} \| \mu'_{T-1}) \eta_\varepsilon(\mathsf{K}_T)$$

$$\leq \mathsf{E}_\varepsilon(\mu_{T-2} \| \mu'_{T-2}) \eta_\varepsilon(\mathsf{K}_T) \eta_\varepsilon(\mathsf{K}_{T-1})$$

⋮

$$\leq \mathsf{E}_\varepsilon(\mu_1 \| \mu'_1) \prod_{t=1}^T \eta_\varepsilon(\mathsf{K}_t) = \mathsf{E}_\varepsilon(\mu_1 \| \mu'_1) \left[\theta_\varepsilon \left(\frac{D}{\sigma} \right) \right]^T$$

□

DP-SGD

Theorem. Assume $\|\Psi_t(w) - \Psi_t(w')\| \leq D$ for all t and w, w' . Then we have

$$\mathsf{E}_\varepsilon(\mu_T \| \mu'_T) \leq \mathsf{E}_\varepsilon(\mu_1 \| \mu'_1) \left[\theta_\varepsilon \left(\frac{D}{\sigma} \right) \right]^T,$$

where $\theta_\varepsilon(r) \triangleq Q\left(\frac{\varepsilon}{r} - \frac{r}{2}\right) - e^\varepsilon Q\left(\frac{\varepsilon}{r} + \frac{r}{2}\right)$ and $Q(a) \triangleq \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-u^2/2} du$.

Projected Noisy SGD:
$$W_{t+1} = \Pi_{\mathcal{K}}(W_t - \eta \nabla \ell(W_t, x_{t+1}) + \eta Z_{t+1})$$

Corollary. Assuming \mathcal{K} is **compact** and $\ell(\cdot, x)$ is L -Lip, the **randomly-stopped** SGD is (ε, δ) -DP for any $\varepsilon > 0$ and

$$\delta = \frac{1}{T} \theta_\varepsilon \left(\frac{2L}{\sigma} \right) \left[1 - \theta_\varepsilon \left(\frac{1 + 2\eta L}{\eta \sigma} \right) \right]^{-1}.$$

DP-SGD

Theorem [Feldman et al.]. Let $\ell(\cdot, x)$ be convex, L -Lip and β -smooth. Then for $\eta \leq \frac{2}{\beta}$ and $\alpha > 1$, the randomly-stopped SGD algorithm is (α, ζ) -RDP where

$$\zeta = \frac{4\alpha L^2 \log T}{T\sigma^2}$$

if $\sigma^2 \geq 2\alpha(\alpha - 1)L^2$.

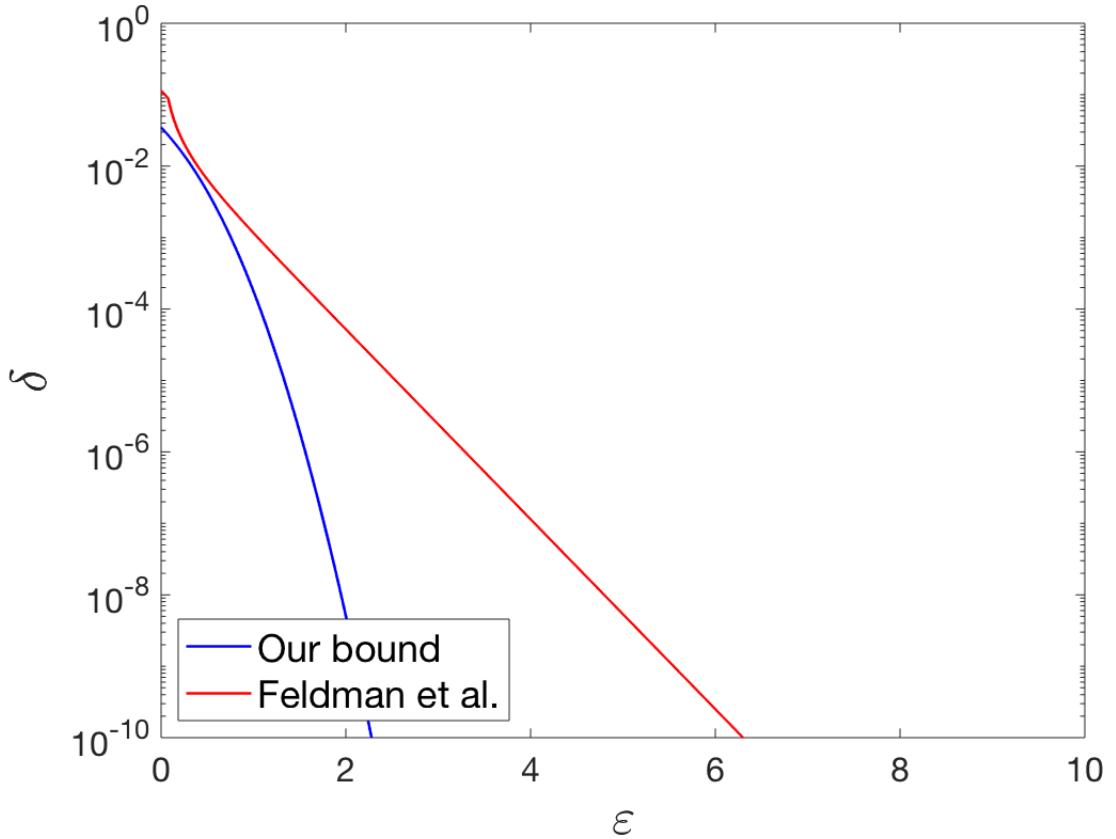
The constraint on σ^2 is due to **non-convexity** of $(P, Q) \mapsto D_\alpha(P \| Q)$ for $\alpha > 1$

Corollary. Given the assumptions above, the randomly-stopped SGD is (ε, δ) -DP for any $\delta > 0$ and

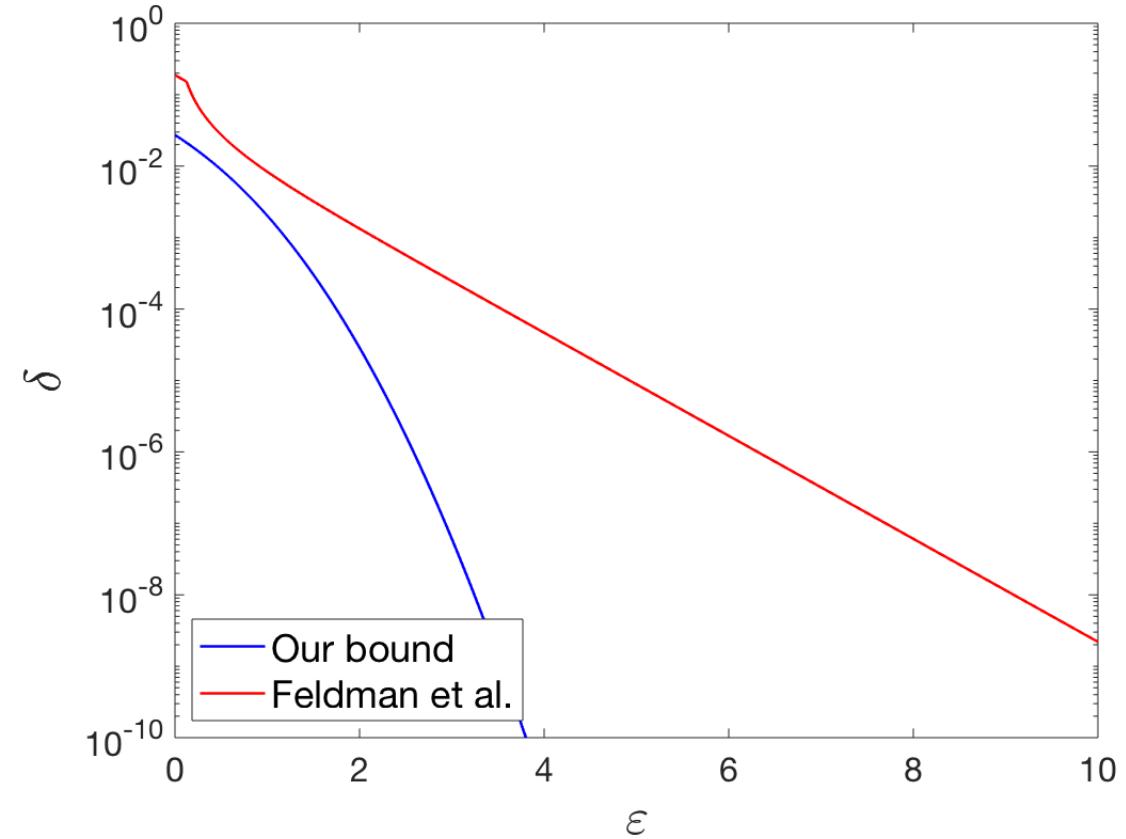
$$\varepsilon = \inf_{\alpha \in (1, \alpha^*]} \frac{1}{\alpha - 1} \min \left\{ (\alpha - 1)\alpha\zeta - \log \frac{\delta}{h_\alpha}, \log \left(\frac{e^{(\alpha-1)\alpha\zeta} - 1}{\alpha\delta} + 1 \right) \right\},$$

where $\alpha^* \triangleq \frac{1}{2} \left[1 + \sqrt{1 + \frac{2\sigma^2}{L^2}} \right]$.

DP-SGD



$$\eta = 0.05, L = 1, T = 100, \sigma = 5$$



$$\eta = 0.1, L = 1, T = 100, \sigma = 3$$

For sufficiently large ε :

- Feldman et al.: $\delta = e^{-O(\varepsilon)}$
- Our result: $\delta = e^{-O(\varepsilon^2)}$

Summary:

- Derive the **lossless** relationship between RDP and DP
- “Improved” moments accountant
- A better bound gives a hundred rounds!
- New privacy analysis of distributed optimization problems via SDPI
- Contraction coefficient of Gaussian kernel
- Tighter bounds than [Feldman et al.’18]

Not discussed:

- Improved moments accountant vs. Gaussian DP
- “Some” operational interpretations of RDP in hypothesis testing
- SDPI-based privacy analysis of general online algorithms

