A Note on Privacy Composition and Amplification

March 23, 2020

This is a note for the paper "Privacy Amplification by iteration"

1 Introduction

1.1 Opitimization Notion

Convex Loss Minimization
$\mathcal{X}$: domain of data sets
$\mathcal{P}$: a distribution over $\mathcal{X}$
$S = \{x_1, ..., x_n\}$: a data set drawn i.i.d. from $\mathcal{P}$
$\mathcal{K}$: a convex set denoting the space of all models, $\mathcal{K} \in \mathbb{R}^d$
$f : \mathcal{K} \times \mathcal{X} \to \mathbb{R}$ is a loss function.
excess population loss of solution: $\mathbb{E}_{x\sim\mathcal{P}}[f(w, x)] - \min_{v\in\mathcal{K}} \mathbb{E}_{x\sim\mathcal{P}}[f(v, x)]$

1.2 Measure Notion

Definition 1.1. ***Measure Absolutely Continuous***
We say a distribution μ is absolutely continuous with respect to ν if $\mu(A) = 0$ whenever $\nu(A) = 0$ for all measurable sets A. We will denote this by $\mu \ll \nu$.

Given tow distributions μ and ν on a Banach space$(\mathcal{Z}, || \cdot ||)$, one can define several notions fo distance between them.

Definition 1.2. ***Rényi Divergence***
Let $1 < \alpha < \infty$ and μ, ν be measures with $\mu \ll \nu$. The Rényi divergence of order α between μ and ν is defined as

$$D_\alpha(u||v) = \frac{1}{\alpha - 1} ln \int (\frac{\mu(z)}{\nu(z)})^\alpha \nu(z) dz$$

It has the following properties:

- It's **independent** of **norm**.
- **Additibity** : $D_\alpha(\mu \times \mu' || \nu \times \nu') = D_\alpha(\mu||\nu) + D_\alpha(\mu'||\nu')$
  Proof: Write p.d.f. for catesian product of measures directy and it's easy to get the result.
- **Post-Processing** : For any(deterministic) function f, $D_\alpha(f(\mu)||f(\nu)) \le D_\alpha(\mu||\nu)$
  Proof: First prove discrete version and Use inversion formula and apply convexity, then approximate continous version using the proof for discrete version.
  Remark: What is reason for deviding it by $\alpha - 1$? To make it approximate entropy when $\alpha \to 1$?

Definition 1.3. ∞***-Wasserstein Distance***
The ∞-Wasserstein distance between distributions μ and ν on a Banach space $(\mathcal{Z}, || \cdot ||)$ is defined as

$$W_\infty(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu,\nu)} \operatorname*{ess\,sup}_{(x,y)\sim\gamma} ||x - y||$$

Jiayuan Ye, Email:yuan74@mail.ustc.edu.cn
School of Mathematics, University of Science and Technology of China, Hefei, Anhui, 230026, China

1.3 Privacy Notion

At a semantic level, we can define (ϵ, δ) differential privacy with regard to neiboring datasets. A common choice if $\epsilon = 0.1, \delta = 1/n^{w(1)}$, where n refers to the size of the dataset. There are times when the traditional approach fails.(PAE+17,PSM+18)

Starting with concentrated Differential Privacy, there has been definitions that allow more fine-grained control of the privacy loss random variable, such as **zCDP,Moments Acountant** and **Rényi differential privacy**.

Definition 1.4. ***[Mir17] Rényi Differential Privacy(RDP).***

For $1 \leq a \leq \infty$ and $\epsilon \geq 0$, a randomized algorithm $\mathcal{A}$ is (α, ϵ)-Rényi differentially private, or $(\alpha, \epsilon) - RDP$ is for all neighboring data sets S and S' we have

$$D_\alpha(A(S)||A(S')) \leq \epsilon$$

Definition 1.5. ***Shifted Rényi Divergence***

Let μ and ν be distributions defined on a Banach space $(\mathcal{Z}, ||\cdot||)$. For parameters $z > 0$ and $\alpha \geq 1$, the z-shifted Rényi divergence between μ and ν is defined as

$$D_\alpha^{(z)}(\mu||\nu) = \inf_{\mu': W_\infty(\mu,\mu') \leq z} D_\alpha(\mu'||\nu)$$

It has the following properties:

- *Monotonicity: for $0 \leq z \leq z', D_\alpha^{(z)}(\mu||\nu) \geq D_\alpha^{(z')}(\mu||\nu)$*
- *Shifting: $D_\alpha^{(||x||)}(\mu||\nu) \leq D_\alpha(\mu * \boldsymbol{x}||\nu)$*

Definition 1.6. *$(R_\alpha(\zeta, a))$*

$$R_\alpha(\zeta, a) = \sup_{x:||x|| \leq a} D_\alpha(\zeta * \boldsymbol{x}||\zeta)$$

Remark:

- $D_\alpha(\mathcal{N}(0, \sigma^2\mathbb{I}_d)||\mathcal{N}(x, \sigma^2\mathbb{I}_d)) = \alpha||x||_2^2/2\sigma^2 \Rightarrow R_\alpha(\mathcal{N}(0, \sigma^2\mathbb{I}_d), x) = \alpha a^2/2\sigma^2$
  Simply write out the p.d.f. then do integration
- *It measures how well noise distribution ζ hids changes in our norm $||\cdot||$*

Lemma 1.1. ***Relating RDP and DP***

If $\mathcal{A}$ satisfies (α, ϵ)-Rényi differential privacy, then for all $\delta \in (0, 1)$, it also satisfies $(\epsilon + \frac{ln(1/\delta)}{\alpha - 1}, \delta)$-differential privacy. Moreover, pure $(\epsilon, 0)$-differential privacy coincides with (∞, ϵ)-RDP.

Proof: Needs to be supplemented

2 Privacy composition

It enables modular design and analysis and controlls the total privacy budget of the combination of simpler building blocks.

Naïve Composition Theorems for DP

Advanced Composition Theorems for DP

An Example(Noisy SGD)

This section needs to be elaborated. Can read on the blog by Rishav Chourasia.

Remark:

- All existing proofs of advanced composition theorems assume that **all intermediate outputs** are revealed, whether the composite mechanism requires it or not.

Jiayuan Ye, Email:yuan74@mail.ustc.edu.cn
School of Mathematics, University of Science and Technology of China, Hefei, Anhui, 230026, China

Lemma 2.1. ***A naive compositon rule for RDP***

If $\mathcal{A}_1, ..., \mathcal{A}_k$ are randomized algorithms satisfying, respectively, $(\alpha, \epsilon_1) - RDP, ..., (\alpha, \epsilon_k) - RDP$,then their composition defined as $(\mathcal{A}_1(S), ..., \mathcal{A}_k(S))$ is $(\alpha, \epsilon_1 + ... + \epsilon_k) - RDP$. Moreover, the i'th algorithm can be chosenon the basis of the outputs of algorithms $\mathcal{A}_1, ..., \mathcal{A}_{i-1}$

Proof: Simple calculation.

Definition 2.1. ***Contractive function***

Proposition 2.2. *For **convex** and β**-smooth** functions, gradient descent function ψ is contractive when $\eta \leq 2/\beta$*

$$\psi(w) = w - \eta \nabla_w f(w)$$

One proof assumed f to be twice differentiable, and then use the mean value theorem. And use the fact that $0 \prec \nabla^2 f(z) \prec \beta \mathbb{I}$

In BBN,

$$l(w) = \sum_{i=1}^{N} log(p(X_i|w)) + log(p(w))$$

Definition 2.2. ***Contractive Noisy Iteration(CNI)***

Given an initial random state $X_0 \in \mathcal{Z}$, a sequence of contractive functions $\psi_t : \mathcal{Z} \to \mathcal{Z}$,and a sequence of noise distributions $\{\zeta_t\}$, we define the Contractive Noisy Iteration(CNI) by the following update rule:

$$X_{t+1} = \psi_{t+1}(X_t) + Z_{t+1}$$

where Z_{t+1} is drawn independently from ζ_{t+1}. We denote the r.v. output by this process after T steps as $CNI_T(X_0, \{\psi_t\}, \{\zeta_t\})$

3 Privacy amplification by sampling

It bounds the privacy budget—for select mechanisms—of a combination to be less than the privacy budget of its parts.

This is the only systematically studied instance of privacy amplification.

4 Privacy amplification by iteration

It constructs a coupled descent to prove a bound on the Rényi divergence, hence on differential privacy.

Main idea:

$$W_\infty(X_0, X_0') \leq 1 \xrightarrow{map's\ contractivity + adding\ noise\ \zeta} X_1\ and\ X_1'\ are\ R_\alpha(\zeta, 1) - close$$

- not applying amplification result

$$\xrightarrow{post-processing\ property\ of\ RD} X_T\ and\ X_T'\ are\ similarly\ close$$

- using amplification result

  Our main theorem says that this can be substantially improved if we do not release the intermediate steps. The noise added in subsequent steps further decreases the Ré nyi divergence even when contractive steps are taken in between the noise addition.

The idea of proving the final result about Rényi divergences is using shifted Rényi divergences as a crucial tool.

Shift reduction lemma: reduce the shift parameter z in the noise addition step.

contractive maps affect the shifted divergence

Jiayuan Ye, Email:yuan74@mail.ustc.edu.cn
School of Mathematics, University of Science and Technology of China, Hefei, Anhui, 230026, China

Lemma 4.1. ***Shift-Reduction Lemma***
Let μ, ν and ζ be distributions over a Banach space $(\mathcal{Z}, ||\cdot||)$. Then for any $a \geq 0$,

$$D_\alpha^{(z)}(\mu * \zeta || \nu * \zeta) \leq D_\alpha^{(z+a)}(\mu||\nu) + R_\alpha(\zeta, a)$$

Proof:
Remark:

- *An intuitive understanding of this lemma is that when adding noise, the difference of the resulting distribution can be controlled by the initial difference and the noise.*
- *Is the difference of the distribution decreasing or increasing with added noise?*
- *The larger the shift, the better bound we get for DP. So shift reduction is not the aim but the tool?*

Lemma 4.2. ***Contraction reduces $D_\alpha^{(z)}$***
Suppose that ψ and ψ' are contractive maps on $(\mathcal{Z}, ||\cdot||)$ and $\sup_x ||\psi(x) - \psi(x')|| \leq s$. Then for r.v.'s X and X' over $\mathcal{Z}$,

$$D_\alpha^{(z+s)}(\psi(X)||\psi'(X')) \leq D_\alpha^{(z)}(X||X')$$

Proof: There is a joint distribution (X, Y) such that $D_\alpha(Y||X') = D_\alpha^{(z)}(X||X)$ and $Pr[||X - Y|| \leq z] = 1$
The key parameter in this lemma is s. It evaluated the difference of the function. In most settings, this s is brought by the difference in the initial data batch.

Theorem 4.3. *Let X_T and $X_{T'}$ denote the output of $CNI_T(X_0, \{\psi_t\}, \{\zeta_t\})$. Let*
Remark: The key parameter in this theorem is a sequence of z_t, s_t and a_t.

- *The root parameter is s_t, which denoted the difference of the gradient descent function for each step.*
- *a_i is what needs to be carefully chosen. It's value determines two things: 1) the increase or decrease of z_t. 2)the privacy loss result that we finally achieve.*
- *z_t is determined by a_t and s_t.*
- *If we replace contractive function by Lipschiz continuous function?*

4.1 Algorithms

Algorithm 4.1. Baseline: PNSGD (Projected Noisy Stochastic gradient descent)

Theorem 4.4. *If parameter space is convex set and objective function is convex, L-Lipschitz and $\beta-smooth$ functions on parameter space,$\eta \leq 2/\beta, \sigma > 0, \alpha > 1$ Then $PNSGD(S, w_0, \eta, \sigma)$ satisfies $(\alpha, \frac{\alpha\epsilon}{n+1-t}) - RDP$ for its t'th input, where $\epsilon = \frac{2L^2}{\sigma^2}$*

Proof:It chose $a_i = \begin{cases} 0 & i \leq t-1 \\ \frac{2\eta L}{n-t+1} & i > t \end{cases}$, $s_t = 2\eta L, s_i = 0$ *for* $i \neq t$

it didn't use the condition $\beta - smooth$. Choosing $_i = \begin{cases} 0 & i \leq t-1 \\ \frac{\eta\beta||X_t - X_t'||}{n-t+1} & i > t \end{cases}$ *and* $s_t = \eta\beta||X_t - X_t'||$ *also seems OK.*

Then apply Theorem 4.3, $D_\alpha(X_n||X_n') \leq \frac{\alpha}{2\eta^2\sigma^2}\sum_{i=1}^{n} a_t^2 \leq \frac{2\alpha L^2}{\sigma^2(n+1-t)}$

Alternative bound is $\frac{\alpha\beta^2||X_t - X_t'||^2}{2\sigma^2(n+1-t)}$

- The batch size is 1?
- why are the a_t chosen this way?

Algorithm 4.2. At least as private: skip-PNSGD

Jiayuan Ye, Email:yuan74@mail.ustc.edu.cn
School of Mathematics, University of Science and Technology of China, Hefei, Anhui, 230026, China

Theorem 4.5. *If parameter space is convex set and objective function is convex, L-Lipschitz and $\beta-smooth$ functions on parameter space.$\eta \leq 2/\beta, \sigma > 0, \alpha > 1, t \in [n]$ Then $PNSGD(S, w_0, \eta, \sigma)$ satisfies $(\alpha, \frac{\alpha\epsilon}{n+1-t}) - RDP$ for its t'th input, where $\epsilon = \frac{2L^2}{\sigma^2}$*

Don't know the aim of skipping. can further read about it but may not be very important?

Algorithm 4.3. random stoping: Stop-PNSGD

Theorem 4.6. *If parameter space is convex set and objective function is convex, L-Lipschitz and $\beta-smooth$ functions on parameter space.$\eta \leq 2/\beta, \alpha > 1, \sigma > L\sqrt{2(\alpha-1)\alpha}$. Then $PNSGD(S, w_0, \eta, \sigma)$ satisfies $(\alpha, \frac{4\alpha L^2 ln\ n}{n\sigma^2}) - RDP$*

Proof: The idea is to use expectation for RDP on the probability space of stopping time to approximate the real RDP. And use the PNSGD theorem to calculated RDP for PNSGD under various stopping time.

The paper said that the skip-PNSGD and Stop-PNSGD satisfies local DP even without the smoothness condition.

5 Applications

Needs to be further elaborated.

6 removing the smoothness requirement

The paper mentioned the method of approximating convex objective function with convex and smooth objective function.

Jiayuan Ye, Email:yuan74@mail.ustc.edu.cn
School of Mathematics, University of Science and Technology of China, Hefei, Anhui, 230026, China