A Note on Privacy Composition and Amplification

March 6, 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: Use inversion formula and Cachy inequalities?

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

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

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

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

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 $||\cdot||$*

Definition 1.7. *[**Mir17**]. 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*

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

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:

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

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

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 $\eta \leq 2/\beta$*

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

Definition 2.2. ***Contractive Noisy Iteration(CNI)***
Needs to be elaborated

3 Privacy amplification

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

3.1 Amplification by sampling

This is the only systematically studied instance of privacy amplification.

3.2 Amplification by iteration

Lemma 3.1. ***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: 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.
Remark:

- *Is the difference of the distribution decreasing or increasing with added noise?*
- *some part of the proof is suspiscious*

Lemma 3.2. ***Contraction reduces D_α^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: Needs to be elaborated.
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.
What I don't understand is why it says contraction reduces $D_\alpha^{(z)}$

Theorem 3.3. *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.*

Algorithm 3.1. ***Projected Noisy Stochastic gradient descent***

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