The Saddle-Point Accountant for Differential Privacy

Shahab Asoodeh

Joint work with Wael Alghamdi, Felipe Gomez, Flavio Calmon (Harvard University), Oliver Kosut, Lalitha Sankar (ASU)

October 19, 2022

#Women-Life-Freedom
Joint work with

Wael Alghamdi (Harvard)

Flavio Calmon (Harvard)

Felipe Gomez (Harvard)

Oliver Kosut (ASU)

Lalitha Sankar (ASU)
State-of-the-art composition

DP-SGD:

\[ \sigma = 0.65 \]

subsampling = \(10^{-2}\)

\[ \delta = 10^{-5} \]
State-of-the-art composition

DP-SGD:
\[ \sigma = 0.65 \]

subsampling = \( 10^{-2} \)
\[ \delta = 10^{-5} \]

runtime complexity
\[ O(\sqrt{n} \log n) \]
State-of-the-art composition

DP-SGD:

\[ \sigma = 0.65 \]

subsampling = \( 10^{-2} \)

\[ \delta = 10^{-5} \]

runtime complexity

\[ O(\sqrt{n} \log n) \]

\[ O(\text{polylog}(n)) \]

---

Gopi, Lee, and Wutschitz, Numerical Composition of Differential Privacy, NeurIPS 2021

Ghazi, Kamath, Kumar, and Manurangsi, Faster Privacy Accounting via Evolving Discretization, ICML 2022
State-of-the-art composition

DP-SGD:

\[ \sigma = 0.65 \]

\[ \text{subsampling} = 10^{-2} \]

\[ n = 2000 \]

Gopi, Lee, and Wutschitz, Numerical Composition of Differential Privacy, NeurIPS 2021
State-of-the-art composition

DP-SGD:

\[ \sigma = 1 \]

subsampling = \(10^{-2}\)

\[ n = 2000 \]
State-of-the-art composition

DP-SGD:

\( \sigma = 1 \)

subsampling = \( 10^{-2} \)

\( n = 2000 \)

Gopi, Lee, and Wutschitz, Numerical Composition of Differential Privacy, NeurIPS 2021
Today’s talk: Develop DP numerical composition using saddle-point approximation:

- Runtime complexity independent of # composition
- Works for all epsilon and delta
Saddle-point vs. state-of-the-art

DP-SGD:
\[ \sigma = 0.65 \]
\[ \text{subsampling} = 10^{-2} \]
\[ \delta = 10^{-5} \]

runtime complexity \( O(1) \)

Gopi, Lee, and Wutschitz, Numerical Composition of Differential Privacy, NeurIPS 2021
Saddle-point vs. state-of-the-art

DP-SGD:

\( \sigma = 1 \)

subsampling = 10\(^{-2}\)

\( n = 2000 \)
Differential privacy

\[ M \text{ is } (\varepsilon, \delta)\text{-DP if } \forall D \sim D' \text{, } \sup_{\text{subset } A} [M_D(A) - \varepsilon M_{D'}(A)] \leq \delta \]

Hockey-stick divergence
\[ \mathbb{E}_{\varepsilon}(M_D \| M_{D'}) \]

\[ M \text{ is } (\varepsilon, \delta)\text{-DP if } \forall D \sim D' \text{, } \mathbb{E}_{\varepsilon}(M_D \| M_{D'}) \leq \delta \]
Dominating distribution, PLRV

A pair \((P, Q)\) is said to dominate \(M\) if

\[
\sup_{D \sim D'} E_\varepsilon(M_D \| M_{D'}) \leq E_\varepsilon(P \| Q)
\]

or tightly dominate \(M\) if equality is achieved for all \(\varepsilon\).

\(\delta(\varepsilon) \triangleq \) smallest \(\delta\) such that \(M\) is \((\varepsilon, \delta)\)-DP

privacy curve

If \((P, Q)\) tightly dominates \(M\), then

\[
\delta(\varepsilon) = E_\varepsilon(P \| Q) = \mathbb{E}\left[\left(1 - e^{\varepsilon - L}\right)_{+}\right]
\]

where

\[
L = \log \frac{dP}{dQ}(X) \quad \text{with} \quad X \sim P
\]
A pair \((P, Q)\) is said to dominate \(M\) if
\[
\sup_{D \sim D'} E_\varepsilon(M_D\|M_{D'}) \leq E_\varepsilon(P\|Q)
\]
or tightly dominate \(M\) if equality is achieved for all \(\varepsilon\).

\(\delta(\varepsilon) \triangleq\) smallest \(\delta\) such that \(M\) is \((\varepsilon, \delta)\)-DP

privacy curve

If \((P, Q)\) tightly dominates \(M\), then
\[
\delta(\varepsilon) \leq E_\varepsilon(P\|Q) = E \left[ \left(1 - e^{\varepsilon - L}\right)_+ \right]
\]
where
\[
L = \log \frac{dP}{dQ}(X) \quad \text{with} \quad X \sim P
\]
Algorithm 1 Differentially private SGD (Outline)

Input: Examples \( \{x_1, \ldots, x_N\} \), loss function \( \mathcal{L}(\theta) = \frac{1}{N} \sum \mathcal{L}(\theta, x_i) \). Parameters: learning rate \( \eta_t \), noise scale \( \sigma \), group size \( L \), gradient norm bound \( C \).

Initialize \( \theta_0 \) randomly

for \( t \in [T] \) do
    Take a random sample \( L_t \) with sampling probability \( L/N \)
    Compute gradient
        For each \( i \in L_t \), compute \( g_t(x_i) \leftarrow \nabla_{\theta_t} \mathcal{L}(\theta_t, x_i) \)
    Clip gradient
        \( \tilde{g}_t(x_i) \leftarrow g_t(x_i) / \max\left(1, \frac{\|g_t(x_i)\|_2}{C}\right) \)
    Add noise
        \( \tilde{g}_t \leftarrow \frac{1}{t} \left( \sum_i \tilde{g}_t(x_i) + \mathcal{N}(0, \sigma^2 C^2 I) \right) \)
    Descent
        \( \theta_{t+1} \leftarrow \theta_t - \eta_t \tilde{g}_t \)

Output \( \theta_T \) and compute the overall privacy cost \((\varepsilon, \delta)\) using a privacy accounting method.
**DP-SGD**

**Algorithm 1** Differentially private SGD (Outline)

**Input:** Examples \( \{x_1, \ldots, x_N\} \), loss function \( L(\theta) = \frac{1}{N} \sum L(\theta, x_i) \). Parameters: learning rate \( \eta_t \), noise scale \( \sigma \), group size \( L \), gradient norm bound \( C \).

**Initialize** \( \theta_0 \) randomly

for \( t \in [T] \) do
  Take a random sample \( L_t \) with sampling probability \( L/N \)
  **Compute gradient**
  For each \( i \in L_t \), compute \( g_t(x_i) \leftarrow \nabla_{\theta_t} L(\theta_t, x_i) \)
  **Clip gradient**
  \( \tilde{g}_t(x_i) \leftarrow g_t(x_i) / \max(1, \frac{\|g_t(x_i)\|_2}{C}) \)
  **Add noise**
  \( \tilde{g}_t \leftarrow \frac{1}{L} (\sum \tilde{g}_t(x_i) + \mathcal{N}(0, \sigma^2 C^2)) \)
  **Descent**
  \( \theta_{t+1} \leftarrow \theta_t - \eta_t \tilde{g}_t \)

**Output** \( \theta_T \) and compute the overall privacy cost \( (\varepsilon, \delta) \) using a privacy accounting method.

Tightly dominating distributions for each iteration:

\[
P = p \mathcal{N}(0, \sigma^2 C^2) + (1 - p) \mathcal{N}(C, \sigma^2 C^2) \quad Q = \mathcal{N}(0, \sigma^2 C^2)
\]
**Algorithm 1** Differentially private SGD (Outline)

**Input:** Examples \( \{x_1, \ldots, x_N\} \), loss function \( \mathcal{L}(\theta) = \frac{1}{N} \sum_i \mathcal{L}(\theta, x_i) \). Parameters: learning rate \( \eta_t \), noise scale \( \sigma \), group size \( L \), gradient norm bound \( C \).

**Initialize** \( \theta_0 \) randomly

for \( t \in [T] \) do

Take a random sample \( L_t \) with sampling probability \( L/N \)

Compute gradient

For each \( i \in L_t \), compute \( \mathbf{g}_t(x_i) \leftarrow \nabla_{\theta_t} \mathcal{L}(\theta_t, x_i) \)

Clip gradient

\( \tilde{\mathbf{g}}_t(x_i) \leftarrow \mathbf{g}_t(x_i) / \max(1, \frac{\|\mathbf{g}_t(x_i)\|_2}{C}) \)

Add noise

\( \bar{\mathbf{g}}_t \leftarrow \frac{1}{L} \left( \sum_i \tilde{\mathbf{g}}_t(x_i) + \mathcal{N}(0, \sigma^2 C^2 I) \right) \)

Descent

\( \theta_{t+1} \leftarrow \theta_t - \eta_t \bar{\mathbf{g}}_t \)

Output \( \theta_T \) and compute the overall privacy cost \( (\varepsilon, \delta) \) using a privacy accounting method.

---

Tightly dominating distributions for each iteration:

\[
P = p \mathcal{N}(0, \sigma^2 C^2) + (1 - p) \mathcal{N}(C, \sigma^2 C^2) \\
Q = \mathcal{N}(0, \sigma^2 C^2)
\]

\[
\delta(\varepsilon) = \mathbb{E}_\varepsilon (P \parallel Q) = \mathbb{E} \left[ (1 - e^{\varepsilon - L})_+ \right]
\]

\[
L = \log \left( 1 - p + p \cdot e^{\frac{C(2X - C)}{2\sigma^2}} \right), \quad X \sim P
\]
Composition of DP

\[ D \xrightarrow{\delta_1(\varepsilon)} M^1 \xrightarrow{} M^1_D \]

\[ D \xrightarrow{\delta_2(\varepsilon)} M^2 \xrightarrow{} M^2_D \]
Composition of DP

\[ M = M^1 \circ M^2 \]

\[ \delta_1(\varepsilon) \quad \delta_2(\varepsilon) \]

\[ M^1 \quad M^2 \]

\[ M^1_D \quad M^2_D \]
Composition of DP

$D$

$M = M^1 \circ M^2$

$M^1$

$\delta_1(\varepsilon)$

$M^1_D$

$M^2$

$\delta_2(\varepsilon)$

$M^2_D$

What is the privacy curve $\delta(\varepsilon)$ of $M$?
What is the privacy curve $\delta(\varepsilon)$ of $M$?
Composition of DP

\[ M = M^1 \circ M^2 \]

\[ D \]

What is the privacy curve \( \delta(\varepsilon) \) of \( M \)?

\((P^1, Q^1)\) tightly dominates \( M^1 \)

\((P^2, Q^2)\) tightly dominates \( M^2 \)
Composition of DP

\[ M = M^1 \circ M^2 \]

\[ D \]

\[ M^1 \]
\[ \delta_1(\varepsilon) \]

\[ M^2 \]
\[ \delta_2(\varepsilon) \]

What is the privacy curve \( \delta(\varepsilon) \) of \( M \)?

\[ (P^1, Q^1) \text{ tightly dominates } M^1 \]

\[ (P^2, Q^2) \text{ tightly dominates } M^2 \]

\[ (P^1 \times P^2, Q^1 \times Q^2) \text{ dominates } M \]

Zhu, Dong, and Wang, Optimal Accounting of Differential Privacy via Characteristic Function, AISTAT 2022
Composition of DP

\[ M = M^1 \circ M^2 \]

\( \delta_1(\varepsilon) \)

\( M^1 \)

\( M_D^1 \)

\( \delta_2(\varepsilon) \)

\( M^2 \)

\( M_D^2 \)

What is the privacy curve \( \delta(\varepsilon) \) of \( M \)?

\((P^1, Q^1)\) tightly dominates \( M^1 \)

\((P^2, Q^2)\) tightly dominates \( M^2 \)

\((P^1 \times P^2, Q^1 \times Q^2)\) dominates \( M \)

\[ \delta(\varepsilon) \leq E_\varepsilon(P^1 \times P^2 || Q^1 \times Q^2) \]

Zhu, Dong, and Wang, Optimal Accounting of Differential Privacy via Characteristic Function, AISTAT 2022
Composition of DP

\[ M = M_1 \circ M_2 \]

\[ \delta_1(\varepsilon) \quad \delta_2(\varepsilon) \]

What is the privacy curve \( \delta(\varepsilon) \) of \( M \)?

\( (P^1, Q^1) \) tightly dominates \( M^1 \)
\( (P^2, Q^2) \) tightly dominates \( M^2 \)

\[ \iff \quad (P^1 \times P^2, Q^1 \times Q^2) \text{ dominates } M \]

\[ \delta(\varepsilon) \leq \mathbb{E}_\varepsilon(P^1 \times P^2 \| Q^1 \times Q^2) = \mathbb{E} \left[ (1 - e^{-(L_1 + L_2)})_+ \right] \]

Zhu, Dong, and Wang, Optimal Accounting of Differential Privacy via Characteristic Function, AISTAT 2022
Composition of DP

\[ M = M^1 \circ M^2 \]

What is the privacy curve \( \delta(\varepsilon) \) of \( M \)?

\[
\begin{align*}
(P^1, Q^1) &\text{ tightly dominates } M^1 \implies (P^1 \times P^2, Q^1 \times Q^2) \text{ dominates } M \\
(P^2, Q^2) &\text{ tightly dominates } M^2 \implies (P^1 \times P^2, Q^1 \times Q^2) \text{ dominates } M
\end{align*}
\]

\[
\delta(\varepsilon) \leq \mathbb{E}_\varepsilon (P^1 \times P^2 \| Q^1 \times Q^2) = \mathbb{E} \left[ (1 - e^{\varepsilon(L_1 + L_2)})^+ \right]
\]

Zhu, Dong, and Wang, Optimal Accounting of Differential Privacy via Characteristic Function, AISTAT 2022
Composition of DP

\( M = M^1 \circ \cdots \circ M^n \)

What is the privacy curve \( \delta(\varepsilon) \) of \( M \)?

\( (P^i, Q^i) \) tightly dominates \( M^i \) \( \implies \) \( (P^1 \times \cdots \times P^n, Q^1 \times \cdots \times Q^n) \) dominates \( M \)

Zhu, Dong, and Wang, Optimal Accounting of Differential Privacy via Characteristic Function, AISTAT 2022
Composition of DP

$M = M^1 \circ \cdots \circ M^n$

$(P^i, Q^i)$ tightly dominates $M^i \implies (P^1 \times \cdots \times P^n, Q^1 \times \cdots \times Q^n)$ dominates $M$

$$\delta(\varepsilon) \leq E_\varepsilon(P^1 \times \cdots \times P^n \| Q^1 \times \cdots \times Q^n)$$

What is the privacy curve $\delta(\varepsilon)$ of $M$?

Zhu, Dong, and Wang, Optimal Accounting of Differential Privacy via Characteristic Function, AISTAT 2022
Composition of DP

\( M = M^1 \circ \cdots \circ M^n \)

\( D \)

\( M^1 \)

\( M^2 \)

\( M^n \)

\( M_D^1 \)

\( M_D^2 \)

\( M_D^n \)

What is the privacy curve \( \delta(\varepsilon) \) of \( M \)?

\((P^i, Q^i)\) tightly dominates \( M^i \) \( \implies (P^1 \times \cdots \times P^n, Q^1 \times \cdots \times Q^n)\) dominates \( M \)

\[ \delta(\varepsilon) \leq \mathbb{E}_\varepsilon(P^1 \times \cdots \times P^n \| Q^1 \times \cdots \times Q^n) = \mathbb{E} \left[ (1 - e^{\varepsilon - (L_1 + \cdots + L_n)})_+ \right] \]

Zhu, Dong, and Wang, Optimal Accounting of Differential Privacy via Characteristic Function, AISTAT 2022
Composition Results

$$
\delta(\varepsilon) \leq \mathbb{E} \left[ (1 - e^{\varepsilon - (L_1 + \cdots + L_n)})_+ \right]
$$
Composition Results

\[ \delta(\varepsilon) \leq \mathbb{E} \left[ (1 - e^{-L_1 - \cdots - L_n})^+ \right] \]

- Moments accountant: [Abadi et al’16], [Mironov’17]
- Central limit theorem: [Dong et al’19], [Sommer et al’19]
- Fast Fourier transform: [Koskela et al’20], [Koskela and Honkela’20], [Koskela et al’21], [Gopi et al’21], [Ghazi et al’22]
- Characteristic function: [Zhu et al’22]
- Piece-wise linearization of HS divergence: [Doroshenko et al’22]
Composition Results

\[ \delta(\varepsilon) \leq \mathbb{E} \left[ \left( 1 - e^{\varepsilon - L_1 - \cdots - L_n} \right)_+ \right] \]

- Moments accountant: [Abadi et al’16], [Mironov’17]
- Central limit theorem: [Dong et al’19], [Sommer et al’19]
- Fast Fourier transform: [Koskela et al’20], [Koskela and Honkela’20], [Koskela et al’21], [Gopi et al’21], [Ghazi et al’22]
- Characteristic function: [Zhu et al’22]
- Piece-wise linearization of HS divergence: [Doroshenko et al’22]
Saddle-point accountant

\[
\delta(\varepsilon) \leq \mathbb{E}\left[(1 - e^{\varepsilon - (L_1 + \cdots + L_n)})_+\right]
\]
Saddle-point accountant

\[ \delta(\varepsilon) \leq \mathbb{E} \left[ (1 - e^{\varepsilon - (L_1 + \cdots + L_n)})_+ \right] \]
Saddle-point accountant

\[ \delta(\varepsilon) \leq \mathbb{E} \left[ (1 - e^{\varepsilon - (L_1 + \cdots + L_n)})_{+} \right] = \int_{\mathbb{R}} (1 - e^{\varepsilon - \ell})_{+} f_L(\ell) d\ell \]
Saddle-point accountant

\[
\delta(\varepsilon) \leq \mathbb{E} \left[ (1 - e^{\varepsilon-(L_1+\ldots+L_n)})^+ \right] = \int_{\mathbb{R}} (1 - e^{\varepsilon-\ell})^+ f_L(\ell) \, d\ell \\
= \int_{\mathbb{R}} (1 - e^{\varepsilon-\ell})^+ e^{-t\ell} e^{t\ell} f_L(\ell) \, d\ell
\]
Saddle-point accountant

\[
\delta(\varepsilon) \leq \mathbb{E} \left[ (1 - e^{\varepsilon - (L_1 + \cdots + L_n)})_+ \right] = \int_{\mathbb{R}} (1 - e^{\varepsilon - \ell})_+ f_L(\ell) d\ell
\]

\[
= \int_{\mathbb{R}} (1 - e^{\varepsilon - \ell})_+ e^{-t\ell} e^{t\ell} f_L(\ell) d\ell
\]

\[
= \mathbb{E}[e^{tL}] \int_{\mathbb{R}} (1 - e^{\varepsilon - \ell})_+ e^{-t\ell} e^{t\ell} f_L(\ell) \frac{d\ell}{\mathbb{E}[e^{tL}]}
\]
Saddle-point accountant

\[
\delta(\varepsilon) \leq \mathbb{E}\left[(1 - e^{\varepsilon-(L_1+\cdots+L_n)})^+\right] = \int_{\mathbb{R}} (1 - e^{\varepsilon-\ell}) f_L(\ell) d\ell
\]

\[
= \int_{\mathbb{R}} (1 - e^{\varepsilon-\ell}) e^{-t\ell} e^{t\ell} f_L(\ell) d\ell
\]

MGF of \( L \)

\[
= \mathbb{E}[e^{tL}] \int_{\mathbb{R}} (1 - e^{\varepsilon-\ell}) e^{-t\ell} e^{t\ell} \frac{f_L(\ell)}{\mathbb{E}[e^{tL}]} d\ell
\]
Saddle-point accountant

\[
\delta(\varepsilon) \leq \mathbb{E} \left[ (1 - e^{\varepsilon - (L_1 + \cdots + L_n)})_+ \right] = \int_{\mathbb{R}} (1 - e^{\varepsilon - \ell}) f_L(\ell) d\ell
\]

\[
= \int_{\mathbb{R}} (1 - e^{\varepsilon - \ell}) e^{-t\ell} e^{t\ell} f_L(\ell) d\ell
\]

MGF of \( L \)

pdf of \( \tilde{L} \) exponentially tilted of \( L \)
Saddle-point accountant

\[
\delta(\varepsilon) \leq \mathbb{E} \left[\left(1 - e^{\varepsilon - (L_1 + \cdots + L_n)}\right)_+\right] = \int_{\mathbb{R}} \left(1 - e^{\varepsilon - \ell}\right)_+ f_L(\ell) d\ell
\]

\[
= \int_{\mathbb{R}} \left(1 - e^{\varepsilon - \ell}\right)_+ e^{-t\ell} e^{t\ell} f_L(\ell) d\ell
\]

\[
= \mathbb{E}[e^{tL}] \int_{\mathbb{R}} \left(1 - e^{\varepsilon - \ell}\right)_+ e^{-t\ell} e^{t\ell} \frac{f_L(\ell)}{\mathbb{E}[e^{tL}]} d\ell
\]

\[
= \mathbb{E}[e^{tL}] \int_{\mathbb{R}} e^{-t\ell} \left(1 - e^{\varepsilon - \ell}\right)_+ f_L(\ell) d\ell
\]
Saddle-point accountant

\[
\delta(\varepsilon) \leq \mathbb{E} \left[ (1 - e^{\varepsilon - (L_1 + \cdots + L_n)})_{+} \right] = \int_{\mathbb{R}} (1 - e^{\varepsilon - \ell})_{+} f_L(\ell) d\ell
\]

\[
= \int_{\mathbb{R}} (1 - e^{\varepsilon - \ell})_{+} e^{-\ell} e^{\ell} f_L(\ell) d\ell
\]

\[
= \mathbb{E}[e^{\ell L}] \int_{\mathbb{R}} (1 - e^{\varepsilon - \ell})_{+} e^{-\ell} e^{\ell} f_L(\ell) \frac{d\ell}{\mathbb{E}[e^{\ell L}]}
\]

\[
= \mathbb{E}[e^{\ell L}] \int_{\mathbb{R}} e^{-\ell} (1 - e^{\varepsilon - \ell})_{+} f_L(\ell) d\ell
\]

Plancherel’s Theorem:

\[
\int_{\mathbb{R}} f(x)g(x)dx = \frac{1}{2\pi} \int_{\mathbb{R}} \mathcal{F}(f)(\omega)\mathcal{F}(g)(\omega)d\omega
\]
Saddle-point accountant

\[
\delta(\varepsilon) \leq \mathbb{E} \left[ (1 - e^{\varepsilon - (L_1 + \ldots + L_n)})^+ \right] = \int_{\mathbb{R}} (1 - e^{\varepsilon - \ell}) e^{-t\ell} e^{t\ell} f_L(\ell) d\ell \\
= \int_{\mathbb{R}} (1 - e^{\varepsilon - \ell}) e^{-t\ell} e^{t\ell} f_L(\ell) d\ell \\
= \mathbb{E}[e^{tL}] \int_{\mathbb{R}} (1 - e^{\varepsilon - \ell}) e^{-t\ell} e^{t\ell} \frac{f_L(\ell)}{\mathbb{E}[e^{tL}]} d\ell \\
= \mathbb{E}[e^{tL}] \int_{\mathbb{R}} e^{-t\ell} (1 - e^{\varepsilon - \ell}) f_L(\ell) d\ell \\
= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t + is)} ds
\]
Saddle-point accountant

\[ \delta(\varepsilon) \leq \mathbb{E}\left[ (1 - e^{\varepsilon - (L_1 + \cdots + L_n)})_+ \right] = \int_{\mathbb{R}} (1 - e^{\varepsilon - \ell})_+ f_L(\ell) d\ell \]

\[ = \int_{\mathbb{R}} (1 - e^{\varepsilon - \ell}) e^{-t\ell} e^{t\ell} f_L(\ell) d\ell \]

\[ = \mathbb{E}[e^{tL}] \int_{\mathbb{R}} (1 - e^{\varepsilon - \ell}) e^{-t\ell} e^{t\ell} \frac{f_L(\ell)}{\mathbb{E}[e^{tL}]} d\ell \]

\[ = \mathbb{E}[e^{tL}] \int_{\mathbb{R}} e^{-t\ell} (1 - e^{\varepsilon - \ell})_+ f_L(\ell) d\ell \]

\[ = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t + is)} ds \]

\[ F_\varepsilon(z) \triangleq K_L(z) - z\varepsilon - \log z - \log(1 + z) \]
Saddle-point accountant

\[
delta(\varepsilon) \leq \mathbb{E} \left[ (1 - e^{\varepsilon - (L_1 + \cdots + L_n)})_+ \right] = \int_{\mathbb{R}} (1 - e^{\varepsilon - \ell})_+ f_L(\ell) d\ell
\]

\[
= \int_{\mathbb{R}} (1 - e^{\varepsilon - \ell}) e^{-t\ell} e^{t\ell} f_L(\ell) d\ell
\]

\[
= \mathbb{E}[e^{tL}] \int_{\mathbb{R}} (1 - e^{\varepsilon - \ell})_+ e^{-t\ell} e^{t\ell} f_L(\ell) d\ell \mathbb{E}[e^{tL}]^{-1}
\]

\[
= \mathbb{E}[e^{tL}] \int_{\mathbb{R}} e^{-t\ell} (1 - e^{\varepsilon - \ell})_+ f_L(\ell) d\ell
\]

\[= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds
\]

cumulant generating function (CGF) of \( L \)

\[
K_L(z) = \log \mathbb{E}[e^{zL}]
\]

\[F_\varepsilon(z) \triangleq K_L(z) - z\varepsilon - \log z - \log(1 + z)
\]
Saddle-point accountant

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} \, ds \]

\[ F_\varepsilon(z) \triangleq K_L(z) - \varepsilon - \log z - \log(1 + z) \]
Saddle-point accountant

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} \, ds \]

\[ F_\varepsilon(z) \triangleq K_L(z) - z\varepsilon - \log z - \log(1 + z) \]

Take \( t \) to be saddle-point of \( F_\varepsilon \):
Saddle-point accountant

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t + is)} \, ds \]

\[ F_\varepsilon(z) \triangleq K_L(z) - z\varepsilon - \log z - \log(1 + z) \]

Take \( t \) to be saddle-point of \( F_\varepsilon \): Unique \( t_* \) satisfying \( F'_\varepsilon(t_*) = 0 \), i.e.,
Saddle-point accountant

$$\delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds$$

$$F_\varepsilon(z) \triangleq K_L(z) - \varepsilon z - \log z - \log(1 + z)$$

Take $t$ to be saddle-point of $F_\varepsilon$: Unique $t_*$ satisfying $F_\varepsilon'(t_*) = 0$, i.e.,

$$K'_L(t_*) = \varepsilon + \frac{1}{t_*} + \frac{1}{t_* + 1}$$
Saddle-point accountant

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds \]

\[ F_\varepsilon(z) \triangleq K_L(z) - \varepsilon \log z - \log(1 + z) \]

Take \( t \) to be saddle-point of \( F_\varepsilon \): Unique \( t_* \) satisfying \( F'_\varepsilon(t_*) = 0 \), i.e.,

\[ K_L'(t_*) = \varepsilon + \frac{1}{t_*} + \frac{1}{t_* + 1} \]

\[ F_\varepsilon(z) \approx F_\varepsilon(t_*) + \frac{1}{2}(z - t_*)^2 F''_\varepsilon(t_*) \]
Saddle-point accountant

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_{\varepsilon}(t+i\delta)} ds \]

\[ F_{\varepsilon}(z) \triangleq K_L(z) - \varepsilon - \log z - \log(1 + z) \]

Take \( t \) to be saddle-point of \( F_{\varepsilon} \): Unique \( t_* \) satisfying \( F_{\varepsilon}'(t_*) = 0 \), i.e.,

\[ K'_L(t_*) = \varepsilon + \frac{1}{t_*} + \frac{1}{t_* + 1} \]

Along the real line, \( F_{\varepsilon} \) is minimized at \( z = t_* \)

\[ F_{\varepsilon}(z) \approx F_{\varepsilon}(t_*) + \frac{1}{2}(z - t_*)^2 F_{\varepsilon}'''(t_*) \]
Saddle-point accountant

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds \]

\[ F_\varepsilon(z) \triangleq K_L(z) - \varepsilon \log z - \log(1 + z) \]

Take \( t \) to be saddle-point of \( F_\varepsilon \): Unique \( t_* \) satisfying \( F_\varepsilon'(t_*) = 0 \), i.e.,

\[ F_\varepsilon(z) \approx F_\varepsilon(t_*) + \frac{1}{2}(z - t_*)^2 F_\varepsilon''(t_*) \]

Along the real line, \( F_\varepsilon \) is minimized at \( z = t_* \)

\[ K_L'(t_*) = \varepsilon + \frac{1}{t_*} + \frac{1}{t_* + 1} \]

Parallel to imaginary axis, \( F_\varepsilon \) is maximized at \( z = t_* \).
Saddle-point accountant

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds \]

\[ F_\varepsilon(z) \triangleq K_L(z) - z\varepsilon - \log z - \log(1 + z) \]

Take \( t \) to be saddle-point of \( F_\varepsilon \): Unique \( t_* \) satisfying \( F'_\varepsilon(t_*) = 0 \), i.e.,

\[ K'_L(t_*) = \varepsilon + \frac{1}{t_*} + \frac{1}{t_* + 1} \]
Saddle-point accountant

$$\delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t + is)} ds$$

$$F_\varepsilon(z) \triangleq K_L(z) - z\varepsilon - \log z - \log(1 + z)$$

Take $t$ to be saddle-point of $F_\varepsilon$: Unique $t_*$ satisfying $F_\varepsilon'(t_*) = 0$, i.e.,

First approximation:

Second approximation:
Saddle-point accountant

$$\delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F(\varepsilon(t+i\sigma))} ds$$

$$F(\varepsilon(z) \triangleq K_L(z) - \varepsilon \log z - \log(1 + z)$$

Take $t$ to be saddle-point of $F(\varepsilon)$: Unique $t_*$ satisfying $F'(\varepsilon(t_*)) = 0$, i.e.,

$$K'_L(t_*) = \varepsilon + \frac{1}{t_*} + \frac{1}{t_* + 1}$$

First approximation: Vanilla saddle-point approximation

$$F(\varepsilon(z) \approx F(\varepsilon(t_*)) + \frac{1}{2} (z - t_*)^2 F''(\varepsilon(t_*))$$

Second approximation:
Saddle-point accountant

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds \]

\[ F_\varepsilon(z) \triangleq K_L(z) - \varepsilon z - \log z - \log(1 + z) \]

Take \( t \) to be saddle-point of \( F_\varepsilon \): Unique \( t_* \) satisfying \( F'_\varepsilon(t_*) = 0 \), i.e.,

\[ K'_L(t_*) = \varepsilon + \frac{1}{t_*} + \frac{1}{t_* + 1} \]

First approximation: Vanilla saddle-point approximation

\[ F_\varepsilon(z) \approx F_\varepsilon(t_*) + \frac{1}{2}(z - t_*)^2 F''_\varepsilon(t_*) \]

Second approximation:

\[ K_L(z) \approx K_L(t_*) + \frac{1}{2}(z - t_*)^2 K''_L(t_*) \]
Saddle-point accountant

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+i\varepsilon)} ds \]

Take \( t \) to be saddle-point of \( F_\varepsilon \):
Unique \( t_* \) satisfying \( F_\varepsilon'(t_*) = 0 \), i.e.,

First approximation:

\[ F_\varepsilon(z) \approx F_\varepsilon(t_*) + \frac{1}{2} (z - t_*)^2 F_\varepsilon''(t_*) \]

Second approximation:

\[ K_L(z) \approx K_L(t_*) + \frac{1}{2} (z - t_*)^2 K_L''(t_*) \]
Saddle-point accountant

\[
\delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds
\]

\[
F_\varepsilon(z) \triangleq K_L(z) - z\varepsilon - \log z - \log(1+z)
\]

Take \( t \) to be saddle-point of \( F_\varepsilon \): Unique \( t_* \) satisfying \( F'_\varepsilon(t_*) = 0 \), i.e.,

\[
K'_L(t_*) = \varepsilon + \frac{1}{t_*} + \frac{1}{t_* + 1}
\]

First approximation:

\[
F_\varepsilon(z) \approx F_\varepsilon(t_*) + \frac{1}{2}(z - t_*)^2 F''_\varepsilon(t_*)
\]

\[
\delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds \approx \frac{e^{F_\varepsilon(t_*)}}{\sqrt{2\pi F''_\varepsilon(t_*)}}
\]

Second approximation:

\[
K_L(z) \approx K_L(t_*) + \frac{1}{2}(z - t_*)^2 K''_L(t_*)
\]
Saddle-point accountant

\[
\delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds \\
F_\varepsilon(z) \triangleq K_L(z) - z\varepsilon - \log z - \log(1 + z)
\]

Take \( t \) to be saddle-point of \( F_\varepsilon \): Unique \( t_* \) satisfying \( F_\varepsilon'(t_*) = 0 \), i.e.,

\[
K_L'(t_*) = \varepsilon + \frac{1}{t_*} + \frac{1}{t_* + 1}
\]

First approximation:

\[
F_\varepsilon(z) \approx F_\varepsilon(t_*) + \frac{1}{2}(z - t_*)^2 F_\varepsilon''(t_*)
\]

\[
\downarrow
\]

\[
\delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds \approx \frac{e^{F_\varepsilon(t_*)}}{\sqrt{2\pi F_\varepsilon''(t_*)}} = e^{K_L(t_* - \varepsilon t_*)} \sqrt{2\pi \left[t_*^2(1 + t_*)^2 K_L''(t_*) + t_*^2 + (t_* + 1)^2\right]}
\]

Second approximation:

\[
K_L(z) \approx K_L(t_*) + \frac{1}{2}(z - t_*)^2 K_L''(t_*)
\]
Saddle-point accountant

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} \, ds \]

\[ F_\varepsilon(z) \triangleq K_L(z) - z\varepsilon - \log z - \log(1 + z) \]

Take \( t \) to be saddle-point of \( F_\varepsilon \): Unique \( t_\ast \) satisfying \( F'_\varepsilon(t_\ast) = 0 \), i.e.,

First approximation:

\[ F_\varepsilon(z) \approx F_\varepsilon(t_\ast) + \frac{1}{2}(z - t_\ast)^2 F''_\varepsilon(t_\ast) \]

Second approximation:

\[ K'_L(t_\ast) = \varepsilon + \frac{1}{t_\ast} + \frac{1}{t_\ast + 1} \]

\[ K_L(z) \approx K_L(t_\ast) + \frac{1}{2}(z - t_\ast)^2 K''_L(t_\ast) \]
Saddle-point accountant

$$\delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds$$

$$F_\varepsilon(z) \triangleq K_L(z) - z\varepsilon - \log z - \log(1 + z)$$

Take $t$ to be saddle-point of $F_\varepsilon$: Unique $t_*$ satisfying $F'_\varepsilon(t_*) = 0$, i.e.,

$$K'_L(t_*) = \varepsilon + \frac{1}{t_*} + \frac{1}{t_* + 1}$$

First approximation:

$$F_\varepsilon(z) \approx F_\varepsilon(t_*) + \frac{1}{2}(z - t_*)^2 F''_\varepsilon(t_*)$$

Second approximation:

$$K_L(z) \approx K_L(t_*) + \frac{1}{2}(z - t_*)^2 K''_L(t_*)$$
Saddle-point accountant

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_{\varepsilon}(t+is)} \, ds \]

\[ F_{\varepsilon}(z) \triangleq K_L(z) - z\varepsilon - \log z - \log(1 + z) \]

Take \( t \) to be saddle-point of \( F_{\varepsilon} \): Unique \( t_\ast \) satisfying \( F_{\varepsilon}'(t_\ast) = 0 \), i.e.,

First approximation:

\[ F_{\varepsilon}(z) \approx F_{\varepsilon}(t_\ast) + \frac{1}{2} (z - t_\ast)^2 F_{\varepsilon}''(t_\ast) \]

Second approximation:

\[ K_L(z) \approx K_L(t_\ast) + \frac{1}{2} (z - t_\ast)^2 K_L''(t_\ast) \]

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_{\varepsilon}(t+is)} \, ds \approx e^{K_L(t_\ast) - \varepsilon t_\ast} \mathbb{E} \left[ e^{t_\ast(\varepsilon-Z)} (1 - e^{\varepsilon-Z})^+ \right] \]

\[ K_L'(t_\ast) = \varepsilon + \frac{1}{t_\ast} + \frac{1}{t_\ast + 1} \]
Saddle-point accountant

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds \]

Take \( t \) to be saddle-point of \( F_\varepsilon \): Unique \( t_* \) satisfying \( F_\varepsilon'(t_*) = 0 \), i.e.,

First approximation:

\[ F_\varepsilon(z) \approx F_\varepsilon(t_*) + \frac{1}{2}(z - t_*)^2 F_\varepsilon''(t_*) \]

Second approximation:

\[ K_L(z) \approx K_L(t_*) + \frac{1}{2}(z - t_*)^2 K_L''(t_*) \]

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds \approx e^{K_L(t_*) - \varepsilon t_*} \mathbb{E}\left[ e^{t_*(\varepsilon-Z)}(1-e^{\varepsilon-Z}) \right] \]

\[ Z \sim \mathcal{N}\left( K_L'(t_*), K_L''(t_*) \right) \]
Saddle-point accountant: Algorithm

**Input:** Tightly dominating pairs \{\((P_i, Q_i)\)\}_{i=1}^n\) for mechanisms \{M_i\}_{i=1}^n\) and \(\varepsilon\)

- Compute (numerically estimate) \(K_{L_i}(t) = \log \mathbb{E}[e^{tL_i}]\)

- \(K_L(t) = \sum_{i} K_{L_i}(t)\) (since \(L_i\) are independent)

- Find saddle-point \(t_*\) by solving \(K'_L(t_*) = \varepsilon + \frac{1}{t_*} + \frac{1}{t_*+1}\)

**Outputs:**
Saddle-point accountant: Algorithm

**Input:** Tightly dominating pairs \{\((P_i, Q_i)\)\}_{i=1}^n \text{ for mechanisms } \{M_i\}_{i=1}^n \text{ and } \varepsilon

- Compute (numerically estimate) \(K_{L_i}(t) = \log \mathbb{E}[e^{tL_i}]\)
- \(K_L(t) = \sum_i K_{L_i}(t) \quad \text{(since } L_i \text{ are independent)}\)
- Find saddle-point \(t_*\) by solving \(K'_L(t_*) = \varepsilon + \frac{1}{t_*} + \frac{1}{t_*+1}\)

**Outputs:**

\[
\hat{\delta}_1(\varepsilon) = \frac{e^{K_L(t_*)-\varepsilon t_*}}{\sqrt{2\pi} \left[t_*(1+t_*)^2K''_L(t_*) + t_*^2 + (t_*+1)^2\right]^{1/2}}
\]

\[
\hat{\delta}_2(\varepsilon) = e^{K_L(t_*)-\varepsilon t_*} \mathbb{E}\left[e^{t_*(\varepsilon-Z)}(1-e^{\varepsilon-Z})_+\right]
\]
Saddle-point accountant: Algorithm

**Input:** Tightly dominating pairs \(\{(P_i, Q_i)\}_{i=1}^n\) for mechanisms \(\{M_i\}_{i=1}^n\) and \(\varepsilon\)

- Compute (numerically estimate) \(K_{L_i}(t) = \log \mathbb{E}[e^{tL_i}]\)
- \(K_L(t) = \sum_i K_{L_i}(t)\) (since \(L_i\) are independent)
- Find saddle-point \(t^*_\) by solving \(K'_L(t^*_\) = \(\varepsilon + \frac{1}{t^*_\} + \frac{1}{t^*_+1}\)

**Outputs:**

\[
\hat{\delta}_1(\varepsilon) = \frac{e^{K_L(t^*_\) - \varepsilon t^*_\}}{\sqrt{2\pi \left[ t^*_2(1 + t^*_)^2K''_L(t^*_\) + t^*_2 + (t^*_ + 1)^2 \right]}}
\]

\[
\hat{\delta}_2(\varepsilon) = e^{K_L(t^*_\) - \varepsilon t^*_\} \mathbb{E}\left[ e^{t^*_\(\varepsilon-Z\)}(1 - e^{\varepsilon-Z})_+ \right]
\]

Both are “corrected” versions of moments accountant
Numerical experiments

DP-SGD:

\[ \sigma = 0.65 \]

\[ \text{subsampling} = 10^{-2} \]

\[ \delta = 10^{-5} \]
Numerical experiments

DP-SGD:
\[ \sigma = 0.65 \]
\[ \text{subsampling} = 10^{-2} \]
\[ \delta = 10^{-5} \]

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} \, ds \]

\[ F_\varepsilon(z) \triangleq K_L(z) - z\varepsilon - \log z - \log(1+z) \]
Numerical experiments

DP-SGD:

\[ \sigma = 1 \]

subsampling = \(10^{-2}\)

\[ n = 2000 \]
Error analysis

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds \]

First approximation:

\[ F_\varepsilon(z) \approx F_\varepsilon(t_*) + \frac{1}{2}(z - t_*)^2 F_\varepsilon''(t_*) \]

Second approximation:

\[ K_L(z) \approx K_L(t_*) + \frac{1}{2}(z - t_*)^2 K_L''(t_*) \]
Error analysis

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds \]

First approximation:

\[ F_\varepsilon(z) \approx F_\varepsilon(t_\ast) + \frac{1}{2}(z - t_\ast)^2 F''_\varepsilon(t_\ast) \]

vanilla saddle-point approximation

Second approximation:

\[ K_L(z) \approx K_L(t_\ast) + \frac{1}{2}(z - t_\ast)^2 K''_L(t_\ast) \]

Edgeworth expansion
Error analysis

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds \]

First approximation:
\[ F_\varepsilon(z) \approx F_\varepsilon(t_\ast) + \frac{1}{2} (z - t_\ast)^2 F''_\varepsilon(t_\ast) \]

Second approximation:
\[ K_L(z) \approx K_L(t_\ast) + \frac{1}{2} (z - t_\ast)^2 K''_L(t_\ast) \]

For any \( \varepsilon \geq 0 \)
\[ \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds - \hat{\delta}_2(\varepsilon) \right| \leq e^{K_L(t_\ast) - \varepsilon t_\ast} \left( \frac{t_\ast}{1 + t_\ast} \right)^{t_\ast} \frac{P_{t_\ast}}{K''_L(t_\ast)^{3/2}} \]
Error analysis

\[
\delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+i\varepsilon)} ds
\]

First approximation:

\[
F_\varepsilon(z) \approx F_\varepsilon(t_\ast) + \frac{1}{2}(z - t_\ast)^2 F_\varepsilon''(t_\ast)
\]

Second approximation:

\[
K_L(z) \approx K_L(t_\ast) + \frac{1}{2}(z - t_\ast)^2 K_L''(t_\ast)
\]

For any \( \varepsilon \geq 0 \)

\[
\left| \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+i\varepsilon)} ds - \delta_2(\varepsilon) \right| \leq e^{K_L(t_\ast) - \varepsilon t_\ast} \left( \frac{t_\ast}{1 + t_\ast} \right)^{t_\ast} \frac{P_{t_\ast}}{K_L''(t_\ast)^{3/2}}
\]

\[
\sum_{i=1}^{n} \mathbb{E}[|\bar{L}_i - \mathbb{E}[\bar{L}_i]|^3]
\]
Error analysis

\[ \delta(\varepsilon) \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds \]

First approximation:

\[ F_\varepsilon(z) \approx F_\varepsilon(t_*) + \frac{1}{2}(z - t_*)^2 F_\varepsilon''(t_*) \]

Second approximation:

\[ K_L(z) \approx K_L(t_*) + \frac{1}{2}(z - t_*)^2 K_L''(t_*) \]

For any \( \varepsilon = \mathbb{E}[L] + b \cdot \text{var}(L) \), we have

\[ \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{F_\varepsilon(t+is)} ds - \hat{\delta}_2(\varepsilon) \right| \leq \frac{C}{\sqrt{n}} \]
Error analysis
Summary

• Accountant algorithm comparable with state-of-the-art

• Runtime complexity independent of # composition

• Works for all epsilon and delta
Summary

- Accountant algorithm comparable with state-of-the-art
- Runtime complexity independent of # composition
- Works for all epsilon and delta

Ongoing work

- Tightly dominating distribution of composed mechanisms
- Tighter error analysis
- Efficient estimator for CGF
Summary

• Accountant algorithm comparable with state-of-the-art
• Runtime complexity independent of # composition
• Works for all epsilon and delta

Ongoing work

• Tightly dominating distribution of composed mechanisms
• Tighter error analysis
• Efficient estimator for CGF

Pre-print available here!