Vadym Doroshenko, Badih Ghazi, Pritish Kamath, Ravi Kumar, and Pasin Manurangsi

Connect the Dots: Tighter Discrete Approximations of Privacy Loss Distributions

Abstract: The privacy loss distribution (PLD) provides a tight characterization of the privacy loss of a mechanism in the context of differential privacy (DP). Recent work [18–20, 24] has shown that PLD-based accounting allows for tighter (ε, δ)-DP guarantees for many popular mechanisms compared to other known methods. A key question in PLD-based accounting is how to approximate any (potentially continuous) PLD with a PLD over any specified discrete support.

We present a novel approach to this problem. Our approach supports both *pessimistic* estimation, which overestimates the hockey-stick divergence (i.e., δ) for any value of ε, and *optimistic* estimation, which underestimates the hockey-stick divergence. Moreover, we show that our pessimistic estimate is the *best* possible among all pessimistic estimates. Experimental evaluation shows that our approach can work with much larger discretization intervals while keeping a similar error bound compared to previous approaches and yet give a better approximation than an existing method [24].

Keywords: privacy loss distribution, pessimistic approximation, optimistic approximation, privacy accounting, composition

DOI 10.56553/popets-2022-0122
Received 2022-02-28; revised 2022-06-15; accepted 2022-06-16.

1 Introduction

Differential privacy (DP) [8, 9] has become widely adopted as a notion of privacy in analytics and machine learning applications, leading to numerous practical deployments including in industry [3, 7, 13, 16, 29] and government agencies [2]. The DP guarantee of a (randomized) algorithm is parameterized by two real numbers $\varepsilon > 0$ and $\delta \in [0, 1]$; the smaller these values, the more private the algorithm.

The appeal of DP stems from the strong privacy that it guarantees (which holds even if the adversary controls the inputs of all other users in the database), and from its nice mathematical properties. These include *composition*, whose *basic* form [8] says that executing an $(\varepsilon_1, \delta_1)$-DP algorithm and an $(\varepsilon_2, \delta_2)$-DP algorithm and returning their results gives an algorithm that is $(\varepsilon_1+\varepsilon_2, \delta_1+\delta_2)$-DP. While basic composition can be used to bound the DP properties of k algorithms, it is known to not be tight, in particular for large values of k. In fact, *advanced* composition [12] yields a general improvement, often translating to $\approx \sqrt{k}$ reduction in the ε privacy bound of the composition of k mechanisms each of which being $(\varepsilon_0, \delta_0)$-DP. Such a reduction can be sizeable in practical deployments, and therefore much research has been focusing on obtaining tighter composition bounds in various settings.

In the aforementioned setting where each mechanism has the same DP parameters, Kairouz et al. [17] derived the optimal composition bound. For the more general case of composing k mechanisms whose privacy parameters are possibly different, i.e., the ith mechanism is guaranteed to be $(\varepsilon_i, \delta_i)$-DP for some parameters ε_i, δ_i, computing the (exact) DP parameters of the composed mechanism is known to be #P-complete [27].

While the results of [17, 27] provide a complete picture of privacy accounting when we assume only that the ith mechanism is $(\varepsilon_i, \delta_i)$-DP, we can often arrive at tighter bounds when taking into account some additional information about the privacy loss of the mechanisms. For example, the Moments Accountant [1] and Rényi DP [26] methods keep track of (upper bounds on) the Renyi divergences of the output distributions on two adjacent databases; this allows one to compute upper bounds on the privacy parameters. These tools were originally introduced in the context of deep learning with DP (where the composition is over multiple iterations of the learning algorithm) in which they provide significant improvements over simply using the DP parameters of each mechanism. Other known tools that can also be used to upper-bound the privacy pa-

Vadym Doroshenko: Google, E-mail: dvadym@google.com
Badih Ghazi: Google Research, E-mail: badihghazi@gmail.com
Pritish Kamath: Google Research, E-mail: pritish@alum.mit.edu
Ravi Kumar: Google Research, E-mail: ravi.k53@gmail.com
Pasin Manurangsi: Google Research, E-mail: pasin@google.com

rameters of composed mechanisms include concentrated DP [6, 11] and its truncated variant [5]. These methods are however all known not to be tight, and do not allow a high-accuracy estimation of the privacy parameters.

A numerical method for estimating the privacy parameters of a DP mechanism to an arbitrary accuracy, which has been the subject of several recent works starting with [24, 30], relies on the *privacy loss distribution* (PLD). This is the probability mass function of the so-called privacy loss random variable in the case of discrete mechanisms, and its probability density function in the case of continuous mechanisms. From the PLD of a mechanism, one can easily obtain its (tight) privacy parameters. Moreover, a crucial property is that the PLD of a composition of multiple mechanisms is the convolution of their individual PLDs. Thus, [19] used the Fast Fourier Transform (FFT) in order to speed up the computation of the PLD of the composition. Furthermore, explicit bounds on the approximation error for the resulting algorithm were derived in [15, 18–20]. The PLD has been the basis of multiple open-source implementations from both industry and academia including [22, 23, 25]. We note that the PLD can be applied to mechanisms whose privacy loss random variables do not have bounded moments, and thus for which composition cannot be analyzed using the Moments Accountant or Rényi DP methods. An example such mechanism is DP-SGD-JL from [4].

A crucial step in previous papers that use PLDs is in approximating the distribution so that it has finite support; this is especially needed in the case where the PLD is continuous or has a support of a very large size, as otherwise the FFT cannot be performed efficiently. With the exception of [15][1], previous PLD-based accounting approaches [18–20, 24] employ *pessimistic estimators* and *optimistic estimators* of PLDs. Roughly speaking, the former overestimate (i.e., give upper bounds on) δ, whereas the latter underestimate δ. For efficiency reasons, we would like the support of the approximate PLDs to be as small as possible, while retaining the accuracy of the estimates.

Our Contributions

Our main contributions are the following:

- ▷ We obtain a new pessimistic estimator for a PLD and a given desired support set. Our pessimistic estimator is simple to construct and is based on the idea of "connecting the dots" of the hockey-stick curve at the discretization intervals. Interestingly, we show that this is the best possible pessimistic estimator (and therefore is at least as good as previous estimators).
- ▷ We complement the above result by obtaining a new optimistic estimator that underestimates the PLD. This estimator is based on the combination of a greedy algorithm and a convex hull computation. In contrast to the pessimistic case, we prove that there is no "best" possible optimistic estimator.
- ▷ We conduct an experimental evaluation showing that our estimators can work with much larger discretization intervals while keeping a similar error bound compared to previous approaches and yet give a better approximation than existing methods.

2 Preliminaries

For $k \in \mathbb{N}$, we use $[k]$ to denote $\{1, \ldots, k\}$. For a set $S \subseteq \mathbb{R} \cup \{-\infty, +\infty\}$, we write $\exp(S)$ to denote $\{e^a \mid a \in S\}$. Similarly, for $S \subseteq \mathbb{R}_{\geq 0} \cup \{+\infty\}$, we use $\log(S)$ to denote $\{\log(a) \mid a \in S\}$. Here we use the (standard) convention that $e^{+\infty} = +\infty$ and $e^{-\infty} = 0$; we also use the convention that $(+\infty) \cdot 0 = (-\infty) \cdot 0 = 0$. Moreover, we use $[x]_+$ as a shorthand for $\max\{x, 0\}$.

We use $\mathrm{supp}(P)$ to denote the support of a probability distribution P. For two distributions P, Q, we use $P \otimes Q$ to denote the product distribution of the two. Furthermore, when P, Q are over an additive group, we use $P * Q$ to denote the convolution of the two distributions.

2.1 Hockey-Stick Divergence and Curve

Let $\alpha \geq 0$. The *α-hockey-stick divergence* between two probability distributions P and Q over a domain Ω is given as

$$D_\alpha(P||Q) := \sup_S [P(S) - \alpha \cdot Q(S)]_+ , \qquad (1)$$

where $\sup_S$ is over all measurable sets $S \subseteq \Omega$.

For any pair (A, B) of distributions, let $h_{(A,B)} : \mathbb{R}_{\geq 0} \cup \{+\infty\} \to [0, 1]$ be its *hockey-stick curve*, given as $h_{(A,B)}(\alpha) := D_\alpha(A||B) = \sup_S [A(S) - \alpha \cdot B(S)]_+$.

The following characterization of hockey-stick curves, due to [32], is helpful:

1 Gopi et al. [15] uses an estimator that is neither pessimistic nor optimistic, and instead derive their final values of δ using concentration bound-based error estimates.

Lemma 2.1 ([32]). *A function $h : \mathbb{R}_{\geq 0} \cup \{+\infty\} \to [0, 1]$ is a hockey-stick curve for some pair of distributions if and only if the following three conditions hold:*
(i) h is convex and non-increasing,
(ii) $h(0) = 1$,
(iii) $h(\alpha) \geq [1 - \alpha]_+$ for all $\alpha \in \mathbb{R}_{\geq 0} \cup \{+\infty\}$.

2.2 Differential Privacy

The definition of differential privacy (DP) [8, 9][2] can be stated in terms of the hockey-stick divergence as follows.

Definition 2.2. For a notion of *adjacent datasets*, a mechanism $\mathcal{M}$ is said to satisfy (ε, δ)-*differential privacy* (denoted, (ε, δ)-DP) if for all adjacent datasets $S \sim S'$, it holds that $D_{e^\varepsilon}(\mathcal{M}(S)||\mathcal{M}(S')) \leq \delta$.

We point out that the techniques developed in this paper are general and do not depend on the specific adjacency relation. For the rest of the paper, for convenience, we always use α to denote e^ε.

In most situations however, mechanisms satisfy (ε, δ)-DP for multiple values of ε and δ. This is captured by the *privacy loss profile* $\delta_{\mathcal{M}} : \mathbb{R} \to \mathbb{R}$ of a mechanism $\mathcal{M}$ given as $\delta_{\mathcal{M}}(\varepsilon) := \sup_{S \sim S'} \mathcal{D}_{e^\varepsilon}(\mathcal{M}(S)||\mathcal{M}(S'))$. It will be more convenient to consider the hockey-stick curve instead of the privacy profile. The only difference is that the hockey-stick curve takes $\alpha = e^\varepsilon$ as parameter instead of ε as in the privacy profile.

2.3 Dominating Pairs

A central notion in our work is that of a *dominating pair* for a mechanism, defined by Zhu et al. [32].

Definition 2.3 (Dominating Pairs [32]). A pair (P, Q) of distributions *dominates* a pair (A, B) of distributions if it holds that

$$\forall \alpha \geq 0 \ : \ D_\alpha(A||B) \ \leq \ D_\alpha(P||Q);$$

we denote this as $(A, B) \preceq (P, Q)$.

A pair (P, Q) of distributions is a *dominating pair for a mechanism* $\mathcal{M}$ if for all adjacent datasets $S \sim S'$, it holds that $(\mathcal{M}(S), \mathcal{M}(S')) \preceq (P, Q)$; we denote this as $\mathcal{M} \preceq (P, Q)$.

A pair (P, Q) of distributions is a *tightly dominating pair* for $\mathcal{M}$ if for every $\alpha \geq 0$, it holds that $D_\alpha(P||Q) = \sup_{S \sim S'} D_\alpha(\mathcal{M}(S)||\mathcal{M}(S'))$.

Note that, by definition, $(P, Q) \succeq (A, B)$ if and only if $h_{(P,Q)}$ is no smaller than $h_{(A,B)}$ pointwise, i.e., $h_{(P,Q)}(\alpha) \geq h_{(A,B)}(\alpha)$ for all $\alpha \in \mathbb{R}_{\geq 0} \cup \{+\infty\}$.

The following result highlights the importance of *dominating pairs*.

Theorem 2.4 ([32]). *If $\mathcal{M} \preceq (P, Q)$ and $\mathcal{M}' \preceq (P', Q')$, then $\mathcal{M} \circ \mathcal{M}' \preceq (P \otimes P', Q \otimes Q')$, where $\mathcal{M} \circ \mathcal{M}'$ is the composition of $\mathcal{M}$ and $\mathcal{M}'$. Furthermore, this holds even for* adaptive *composition.*[3]

Thus, in order to upper bound the privacy loss profile $\delta_{\mathcal{M}}(\varepsilon) := \sup_{S \sim S'} D_{e^\varepsilon}(\mathcal{M}(S)||\mathcal{M}(S'))$, it suffices to compute $D_{e^\varepsilon}(P||Q)$ for a dominating (P, Q) pair for $\mathcal{M}$.

2.4 Privacy Loss Distribution

Privacy Loss Distribution (PLD) [11, 30] is yet another way to represent the privacy loss. For simplicity, we give a definition below specific to discrete distributions P, Q; it can be extended, e.g., to continuous distributions by replacing the probability masses $P(o), Q(o)$ with probability densities of P, Q at o.

Definition 2.5 ([11]). The *privacy loss distribution* (PLD) of a pair (P, Q) of discrete distributions, denoted by $\mathsf{PLD}_{(P,Q)}$, is the distribution of the *privacy loss random variable* L generated by drawing $o \sim P$ and let $L = P(o)/Q(o)$.

As alluded to earlier, PLD can be used to compute the hockey-stick divergence [24, 30] (proof provided in Appendix A for completeness):

Lemma 2.6. *For any pair (P, Q) of discrete distributions and $\varepsilon \in \mathbb{R} \cup \{-\infty, +\infty\}$, we have*

$$D_{e^\varepsilon}(P||Q) := \sum_{\varepsilon' \in \mathrm{supp}(\mathsf{PLD}_{(P,Q)})} [1 - e^{\varepsilon - \varepsilon'}]_+ \cdot \mathsf{PLD}_{(P,Q)}(\varepsilon').$$

Note that the RHS term above depends only on $\mathsf{PLD}_{(P,Q)}$ and not directly on P, Q themselves. For

2 For more background on differential privacy, we refer the reader to the monograph [10].

3 In adaptive composition of $\mathcal{M}' \circ \mathcal{M}$, $\mathcal{M}'$ can also take the output of $\mathcal{M}$ as an auxiliary input. Here the $\mathcal{M}' \preceq (P', Q')$ has to hold for all possible auxiliary input.

convenience, we will abbreviate the RHS term as $D_{e^\varepsilon}(\mathsf{PLD}_{(P,Q)})$.

The main advantage in dealing with PLDs is that composition simply corresponds to convolution of PLDs [24, 30]:

Lemma 2.7. *Let P, Q, P', Q' be discrete distributions. Then we have*

$$\mathsf{PLD}_{(P\otimes P', Q\otimes Q')} = \mathsf{PLD}_{(P,Q)} * \mathsf{PLD}_{(P',Q')}.$$

2.5 Accounting Framework via Dominating Pairs and PLDs

Dominating pairs and PLDs form a powerful set of building blocks to perform privacy accounting. Recall that in privacy accounting, we typically have a mechanism $\mathcal{M} = \mathcal{M}_1 \circ \cdots \circ \mathcal{M}_k$ where each $\mathcal{M}_i$ is a "simple" mechanism (e.g., Laplace or Gaussian mechanisms) and we would like to understand the privacy profile of $\mathcal{M}$.

The approach taken in previous works [18–20, 24] can be summarized as follows.[4]

1. Identify a dominating pair (A_i, B_i) for each $\mathcal{M}_i$.
2. Find a *pessimistic estimate*[5] $(P_i^\uparrow, Q_i^\uparrow) \succeq (A_i, B_i)$ such that $\mathsf{PLD}_{(P_i^\uparrow, Q_i^\uparrow)}$ is supported on a certain set of prespecified values.
3. Compute $\mathsf{PLD}^\uparrow = \mathsf{PLD}_{(P_1^\uparrow, Q_1^\uparrow)} * \cdots * \mathsf{PLD}_{(P_k^\uparrow, Q_k^\uparrow)}$.
4. Compute $\delta^\uparrow(\varepsilon)$ from $\mathsf{PLD}^\uparrow$ using the formula from Lemma 2.6.

By Theorem 2.4 and Lemmas 2.6 and 2.7, we can conclude that $\delta^\uparrow(\varepsilon) \geq \delta_{\mathcal{M}}(\varepsilon)$; in other words, the mechanism $\mathcal{M}$ is $(\varepsilon, \delta^\uparrow(\varepsilon))$-DP as desired.

Note that the reason that one needs $\mathsf{PLD}_{(P_i^\uparrow, Q_i^\uparrow)}$ to have finite support in the second step is so that it can be computed efficiently via the Fast Fourier Transform (FFT). Currently, there is only one approach used in previous works, called *Privacy Buckets* [24]. Roughly speaking, this amounts simply to rounding the PLD up to the nearest point in the specified support set. (See Section 4.1 for a more formal description.)

While the above method gives us an upper bound $\delta^\uparrow(\varepsilon)$ of $\delta_{\mathcal{M}}(\varepsilon)$, there are scenarios where we would like to find a *lower bound* on $\delta_{\mathcal{M}}(\varepsilon)$; for example, this can be helpful in determining how tight our upper bound is. Computing such a lower bound is also possible under the similar framework, except that we need to know a list of tightly dominating pairs $(A_1^*, B_1^*), \ldots, (A_k^*, B_k^*)$ such that there exists an adjacent datasets S, S' for which $D_{e^\varepsilon}(\mathcal{M}(S)||\mathcal{M}(S')) = D_{e^\varepsilon}(A_1^* \otimes \cdots \otimes A_k^* || B_1^* \otimes \cdots \otimes B_k^*)$. If such tightly dominating pairs can be identified, then we can follow the same blueprint as above except we replace a pessimistic estimate with an *optimistic estimate* $(P^\downarrow, Q^\downarrow) \preceq (A_i^*, B_i^*)$. This would indeed gives us a lower bound $\delta^\downarrow(\varepsilon)$ of $\delta_{\mathcal{M}}(\varepsilon)$.

The described framework is illustrated in Figure 1.

3 Finitely-Supported PLDs

As alluded to in the previous section, to take full advantage of FFT, it is important that a PLD is discretized in a way such that its support is finite. For exposition purposes, we will assume that the discretization points include $-\infty$ and $+\infty$. We will use $\mathcal{E}$ for discretization points for the PLD and $\mathcal{A}$ for the corresponding discretization points for the hockey-stick curve:

Assumption 3.1. Let $\mathcal{A} = \{\alpha_0, \ldots, \alpha_k\}$ be any finite subset of $\mathbb{R}_{\geq 0} \cup \{+\infty\}$ such that $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_k = +\infty$, and let $\mathcal{E} = \{\varepsilon_0, \ldots, \varepsilon_k\}$ be such that $\varepsilon_0 = -\infty, \varepsilon_k = +\infty$ and $\varepsilon_i = \log(\alpha_i)$ for all $i \in [k-1]$.

For the remaining of this work, we will operate under the above assumption and we will not state this explicitly for brevity.

Using the characterization in Lemma 2.1, we can also characterize the hockey-stick curve of PLDs whose support is on a prespecified finite set $\mathcal{E}$, stated more precisely in the lemma below. Furthermore, the "inverse" part of the lemma yields an algorithm (Algorithm 1) that can construct A, B given $(h(\alpha_i))_{\alpha_i \in \mathcal{A}}$ which we will use in the sequel.

Lemma 3.2. *A function $h : \mathbb{R}_{\geq 0} \cup \{+\infty\} \to [0,1]$ is a hockey-stick curve for some pair (P, Q) such that $\mathrm{supp}(\mathsf{PLD}_{(P,Q)}) \subseteq \mathcal{E}$ if and only if the following conditions hold:*

(i) h is convex and non-increasing,
(ii) $h(0) = 1$,
(iii) $h(\alpha_i) \geq [1 - \alpha_i]_+$ for all $\alpha_i \in \mathcal{A}$,

4 Note that their results are not phrased in terms of dominating pairs, since the latter is only defined and studied in [32]. Nonetheless, these previous works use similar (but more restricted) notions for "worst case" distributions.

5 We remark that this is slightly inaccurate as the "pessimistic estimate" in previous works may actually not be a valid PLD; please see Section 4.1 for a more detailed explanation.

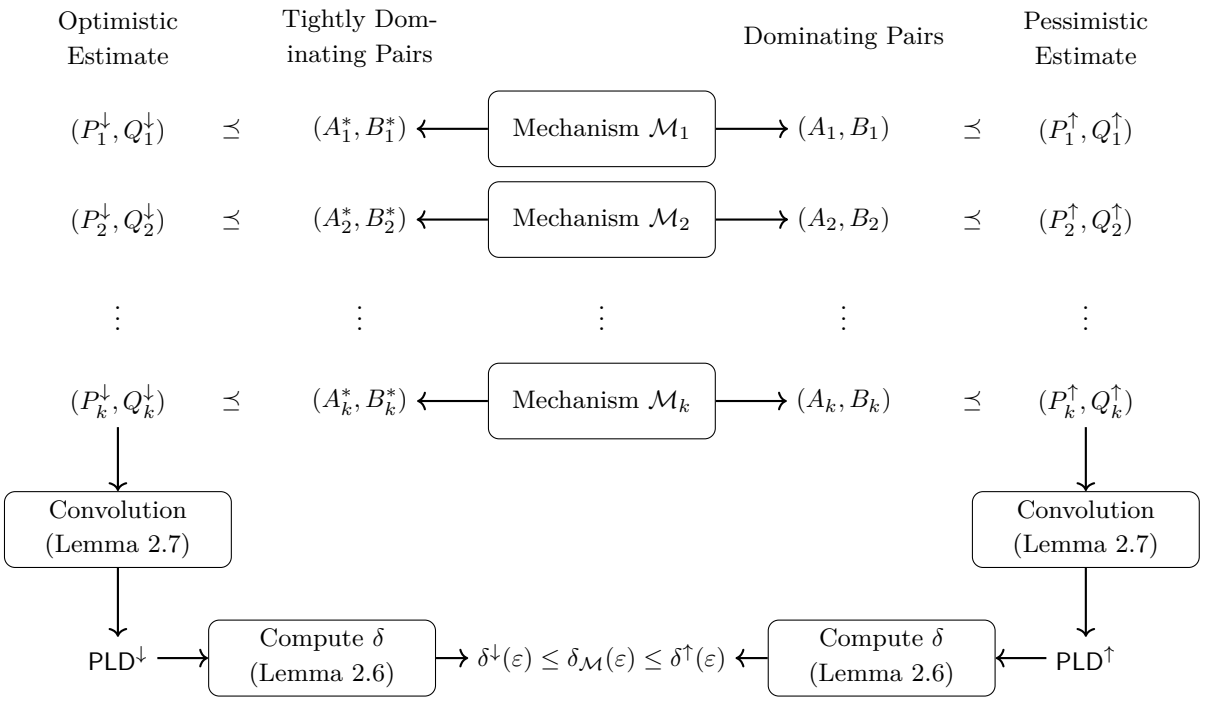


Fig. 1. Illustration of the framework for privacy accounting using PLDs and the notion of (tightly) dominating pairs.

(iv) For all $i \in [k-1]$, the curve h restricted to $[\alpha_{i-1}, \alpha_i]$ is linear: i.e., for all $\alpha \in [\alpha_{i-1}, \alpha_i)$, we have $h(\alpha) = \frac{\alpha-\alpha_i}{\alpha_{i-1}-\alpha_i} \cdot h(\alpha_{i-1}) + \frac{\alpha_{i-1}-\alpha}{\alpha_{i-1}-\alpha_i} \cdot h(\alpha_i)$.
(v) For all $\alpha > \alpha_{k-1}$, $h(\alpha) = h(+\infty)$.

A consequence of Lemma 3.2 is that h is completely specified by $h(\mathcal{A})$. More formally, given $f : \mathcal{A} \to [0,1]$, the only possible extension of f to a hockey-stick curve is its *piecewise-linear* extension $\overline{f}$ defined by

$$\overline{f}(\alpha) := \begin{cases} \frac{\alpha-\alpha_i}{\alpha_{i-1}-\alpha_i} f(\alpha_{i-1}) + \frac{\alpha_{i-1}-\alpha}{\alpha_{i-1}-\alpha_i} f(\alpha_i) & \text{if } \alpha \in [\alpha_{i-1}, \alpha_i) \\ f(+\infty) & \text{if } \alpha > \alpha_{k-1}, \end{cases}$$

for all $\alpha \in \mathbb{R}_{\ge 0} \cup \{+\infty\}$. Note that this $\overline{f}$ may still not be a hockey-stick curve, as it may not be convex.

Proof of Lemma 3.2. ($\Leftarrow$) We start with the converse direction by describing an algorithm that, given $h(\alpha_0), \dots, h(\alpha_k)$, can construct the desired P, Q. In fact, we will construct distributions P and Q with supports contained in $\mathcal{A}$ satisfying the following:

(Π_1) $P(\alpha) = \alpha \cdot Q(\alpha)$ for all $\alpha \in \mathcal{A} \setminus \{+\infty\}$,
(Π_2) $Q(\infty) = 0$,
(Π_3) $D_\alpha(P||Q) = h(\alpha)$ for all $\alpha \in \mathcal{A}$.

The first two conditions imply that $\mathrm{supp}(\mathsf{PLD}_{(P,Q)}) \subseteq \mathcal{E}$ and the last condition implies that $h_{(P,Q)} = h$ as desired.

The construction of P, Q is described in Algorithm 1.

Algorithm 1 PLD Discretization.

procedure DiscretizePLD($h(\alpha_0), \dots, h(\alpha_k)$)
  $Q(\alpha_k) \leftarrow 0$ ▷ $\alpha_k = +\infty$
  for $i = k-1, \dots, 1$ **do**
    $Q(\alpha_i) \leftarrow \frac{h(\alpha_{i-1})-h(\alpha_i)}{\alpha_i-\alpha_{i-1}} - \frac{h(\alpha_i)-h(\alpha_{i+1})}{\alpha_{i+1}-\alpha_i}$
  $Q(\alpha_0) \leftarrow 1 - \sum_{j\in[k-1]} Q(\alpha_j)$ ▷ $\alpha_0 = 0$
  $P(\alpha_0) \leftarrow 0$ ▷ $\alpha_0 = 0$
  for $i = 1, \dots, k-1$ **do**
    $P(\alpha_i) \leftarrow \alpha_i \cdot Q(\alpha_i)$
  $P(\alpha_k) \leftarrow h(\alpha_k)$ ▷ $\alpha_k = +\infty$

Let us now verify that both P and Q are valid probability distributions. First, notice that $Q(\alpha_i) \ge 0$ due to the convexity of h. Furthermore,

$$Q(0) = 1 - \sum_{j\in[k-1]} Q(\alpha_j) = 1 - \frac{1-h(\alpha_1)}{\alpha_1} \ge 0,$$

where the last inequality follows from (iii). Thus, Q is indeed a probability distribution. As for P, notice that $P(\alpha_i) \ge 0$ for all $\alpha_i \in \mathcal{A}$. Finally, we also have

$$\sum_{i\in\{0,\dots,k\}} P(\alpha_i)$$

$$
\begin{aligned}
&= h(+\infty) + \sum_{i\in\{0,\dots,k-1\}} \alpha_i \cdot Q(\alpha_i) \\
&= h(+\infty) + \\
&\quad \sum_{i\in\{0,\dots,k-1\}} \alpha_i \left(\frac{h(\alpha_{i-1}) - h(\alpha_i)}{\alpha_i - \alpha_{i-1}} - \frac{h(\alpha_i) - h(\alpha_{i+1})}{\alpha_{i+1} - \alpha_i} \right) \\
&= h(+\infty) + \sum_{i\in[k-1]} (\alpha_{i+1} - \alpha_i) \cdot \frac{h(\alpha_i) - h(\alpha_{i+1})}{\alpha_{i+1} - \alpha_i} \\
&= h(+\infty) + (h(0) - h(+\infty)) \\
&= 1,
\end{aligned}
$$

meaning that P is a probability distribution as desired.

Properties (Π_1) and (Π_2) are immediate from the construction. We will now check Property (Π_3), based on two cases whether $\alpha > \alpha_{k-1}$.

- ▷ Case I: $\alpha \geq \alpha_{k-1}$. In this case, we have $D_\alpha(P||Q) = P(+\infty) - e^\varepsilon Q(+\infty) = h(+\infty)$.
- ▷ Case II: $\alpha < \alpha_{k-1}$. Suppose that $\alpha \in [\alpha_{i-1}, \alpha_i)$ for $i \in [k-1]$. We have

$$
\begin{aligned}
D_\alpha(P||Q) &= P(\{\alpha_i, \dots, \alpha_k\}) - \alpha \cdot Q(\{\alpha_i, \dots, \alpha_k\}) \\
&= \sum_{j=i}^{k} (\alpha_j - \alpha) \cdot Q(\alpha_j) \\
&= \frac{\alpha - \alpha_i}{\alpha_{i-1} - \alpha_i} \cdot \sum_{j=i}^{k} (\alpha_j - \alpha_{i-1}) \cdot Q(\alpha_j) \\
&\quad + \frac{\alpha_{i-1} - \alpha}{\alpha_{i-1} - \alpha_i} \cdot \sum_{j=i}^{k} (\alpha_j - \alpha_i) \cdot Q(\alpha_j).
\end{aligned}
$$

Furthermore, we have

$$
\begin{aligned}
&\sum_{j=i}^{k} (\alpha_j - \alpha_{i-1}) \cdot Q(\alpha_j) \\
&= \sum_{j=i}^{k} (\alpha_j - \alpha_{i-1}) \\
&\qquad \cdot \left(\frac{h(\alpha_{j-1}) - h(\alpha_j)}{\alpha_j - \alpha_{j-1}} - \frac{h(\alpha_j) - h(\alpha_{j+1})}{\alpha_{j+1} - \alpha_j} \right) \\
&= h(\alpha_{i-1}).
\end{aligned}
$$

Similarly, we also have $\sum_{j=i}^{k} (\alpha_j - \alpha_i) \cdot Q(\alpha_j) = h(\alpha_i)$. Combining the three equalities, we arrive at

$$
D_\alpha(P||Q) = \frac{\alpha - \alpha_i}{\alpha_{i-1} - \alpha_i} \cdot h(\alpha_{i-1}) + \frac{\alpha_{i-1} - \alpha}{\alpha_{i-1} - \alpha_i} \cdot h(\alpha_i),
$$

which is equal to $h(\alpha)$ due to assumption (iv).

As a result, $h_{(P,Q)} = h$ as desired.

($\Rightarrow$) We will now prove this direction. (i), (ii), and (iii) follow immediately from Lemma 2.1. As a result, it suffices to only prove (iv) and (v). Suppose that $h_{(P,Q)} = h$ for some pair (P, Q) such that $\mathrm{supp}(\mathsf{PLD}_{(P,Q)}) \subseteq \mathcal{E}$. Let R be a shorthand for the distribution of $\exp(\mathsf{PLD}_{(P,Q)})$. To prove (iv), consider any $\alpha \in [\alpha_{i-1}, \alpha_i)$ for some $i \in [k-1]$. We then have

$$
\begin{aligned}
h(\alpha) &= D_\alpha(P||Q) \\
&= \sum_{j=i}^{k} (1 - \alpha/\alpha_j) \cdot R(\alpha_j) \\
&= \frac{\alpha - \alpha_i}{\alpha_{i-1} - \alpha_i} \cdot \sum_{j=i}^{k} (1 - \alpha_{i-1}/\alpha_j) \cdot R(\alpha_j) \\
&\quad \frac{\alpha_{i-1} - \alpha}{\alpha_{i-1} - \alpha_i} \cdot \sum_{j=i}^{k} (1 - \alpha_i/\alpha_j) \cdot R(\alpha_j) \\
&= \frac{\alpha - \alpha_i}{\alpha_{i-1} - \alpha_i} \cdot h(\alpha_{i-1}) + \frac{\alpha_{i-1} - \alpha}{\alpha_{i-1} - \alpha_i} \cdot h(\alpha_i),
\end{aligned}
$$

which completes the proof of (iv).

Next, we prove (v). Consider any $\alpha \geq \alpha_{k-1}$. We have

$$
h(\alpha) = D_\alpha(P||Q) = R(+\infty) = h(+\infty),
$$

thereby completing our proof. □

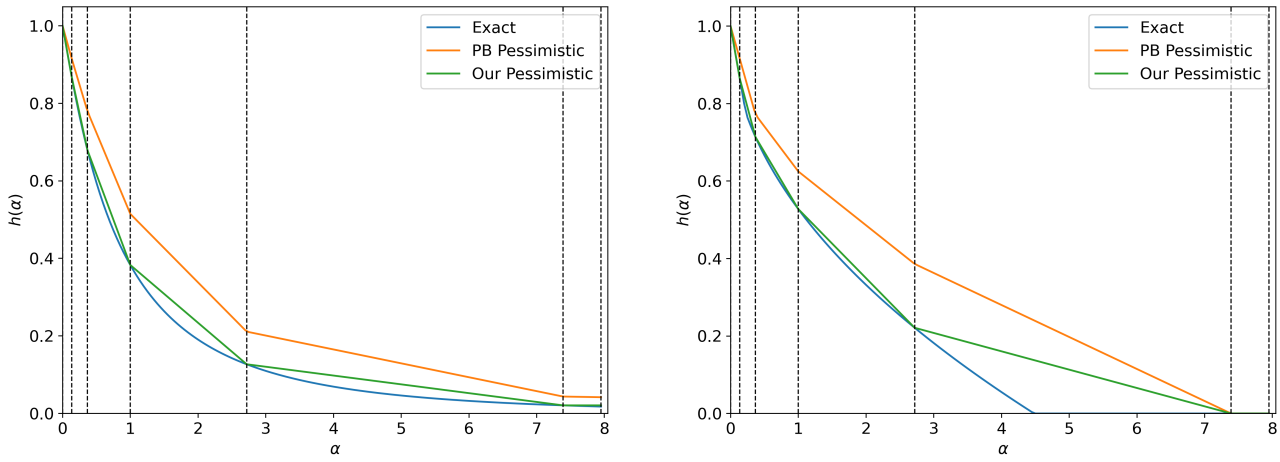
4 Pessimistic PLDs with Finite Support

As we have described in Figure 1, pessimistic estimates of PLDs with finite supports are crucial in the PLD-based privacy accounting framework. The better these pessimistic estimates approximate the true PLD, the more accurate is the resulting upper bound $\delta^\uparrow(\varepsilon)$.

Equipped with tools developed in the previous section, we will now describe our finite-support pessimistic estimate of a PLD. Specifically, given a pair (A, B) of distributions, we would like to compute $(P^\uparrow, Q^\uparrow)$ such that $(P^\uparrow, Q^\uparrow) \succeq (A, B)$ with $\mathrm{supp}(\mathsf{PLD}_{(P^\uparrow,Q^\uparrow)}) \subseteq \mathcal{E}$. In fact, as we will show below (Lemma 4.1), our choice of $\mathsf{PLD}_{(P^\uparrow,Q^\uparrow)}$ "best approximates" $\mathsf{PLD}_{(A,B)}$.

Our construction of the pair $(P^\uparrow, Q^\uparrow)$ is simple: run DiscretizePLD (Algorithm 1) on the input $h_{(A,B)}(\alpha_0), \dots, h_{(A,B)}(\alpha_k)$.

Recall from the proof of Lemma 3.2 that this construction simply gives $h_{(P^\uparrow,Q^\uparrow)}$, which is a piecewise-linear extension of $h_{(A,B)}(\alpha_0), \dots, h_{(A,B)}(\alpha_k)$. In other words, we simply "connect the dots" to construct the hockey-stick curve of our pessimistic estimate. Note that



(a) Gaussian Mechanism

(b) Laplace Mechanism

Fig. 2. Illustrations of the hockey-stick curves of the Gaussian and Laplace mechanisms (with noise multipliers equal to 1 and $2/3$ respectively), and their pessimistic estimates from our approach (labelled "pessimistic") and the Privacy Bucket (PB) approach (labelled "PB pessimistic") of [24]. The horizontal lines represent the discretization points in the set $\mathcal{A}$. As corroborated by Corollary 4.3, our pessimistic estimates is closer to the true curves (labelled "exact") compared to the PB pessimistic estimates for all α.

this, together with the convexity of $h_{(A,B)}$ (Lemma 2.1), implies that $(P^{\uparrow}, Q^{\uparrow}) \succeq (A, B)$ as desired.

Additionally, it is not hard to observe that our choice of pessimistic PLD is the best possible, in sense that $(P^{\uparrow}, Q^{\uparrow})$ is the least element (under the domination partial order) among all pairs that dominate (A, B):

Lemma 4.1. *Let P, Q be any pair of distributions such that $\mathsf{PLD}_{(P,Q)}$ is supported on $\mathcal{A}$ and $(P, Q) \succeq (A, B)$. Then, we must have $(P, Q) \succeq (P^{\uparrow}, Q^{\uparrow})$*

Proof. Recall that it suffices to prove that $h_{(P,Q)}(\alpha) \geq h_{(P^{\uparrow},Q^{\uparrow})}(\alpha)$ for all $\alpha \in \mathbb{R}_{\geq 0} \cup \{+\infty\}$. We will consider two cases based on the value of α:

- ▷ Case I: $\alpha \geq \alpha_{k-1}$. From Lemma 3.2, we simply have $h_{(P,Q)}(\alpha) = h_{(P,Q)}(+\infty) \geq h_{(A,B)}(+\infty) = h_{(P^{\uparrow},Q^{\uparrow})}(\alpha)$.
- ▷ Case II: $\alpha_{k-1} > \alpha \geq 0$. Suppose that $\alpha \in [\alpha_{i-1}, \alpha_i)$. From Lemma 3.2(ii), we have

$$\begin{aligned} & h_{(P,Q)}(\alpha) \\ &= \frac{\alpha - \alpha_i}{\alpha_{i-1} - \alpha_i} \cdot h_{(P,Q)}(\alpha_{i-1}) + \frac{\alpha_{i-1} - \alpha}{\alpha_{i-1} - \alpha_i} \cdot h_{(P,Q)}(\alpha_i) \\ &\geq \frac{\alpha - \alpha_i}{\alpha_{i-1} - \alpha_i} \cdot h(\alpha_{i-1}) + \frac{\alpha_{i-1} - \alpha}{\alpha_{i-1} - \alpha_i} \cdot h(\alpha_i) \\ &= h_{(P^{\uparrow},Q^{\uparrow})}(\alpha), \end{aligned}$$

where the first inequality follows from $(P, Q) \succeq (A, B)$ and the last equality follows from our construction of $(P^{\uparrow}, Q^{\uparrow})$. □

We remark that $\mathsf{PLD}_{(P^{\uparrow},Q^{\uparrow})}$ also has a simple form, due to the properties (Π_1) and (Π_2):

$$\mathsf{PLD}_{(P^{\uparrow},Q^{\uparrow})}(\varepsilon_i) := P^{\uparrow}(\alpha_i)$$

for all $i \in [k]$.

4.1 Comparison to Privacy Loss Buckets

The primary previous work that also derived a pessimistic estimate of PLD is that of Meiser and Mohammadi [24], which has also been used (implicitly) in later works [18–20]. In our terminology, the *Privacy Buckets (PB)* algorithm of Meiser and Mohammadi [24][6] can be restated as follows: let the pessimistic-PB estimate $\widetilde{\mathsf{PLD}}^{\uparrow}_{(A,B)}$ be the probability distribution where

$$\widetilde{\mathsf{PLD}}^{\uparrow}_{(A,B)}(\varepsilon_i) = \mathsf{PLD}_{(A,B)}((\varepsilon_{i-1}, \varepsilon_i]),$$

for all $i \in [k]$. In other words, $\widetilde{\mathsf{PLD}}^{\uparrow}_{(A,B)}$ is a probability distribution on $\mathcal{E}$ that stochastically dominates $\mathsf{PLD}_{(A,B)}$; furthermore, $\widetilde{\mathsf{PLD}}^{\uparrow}_{(A,B)}$ is the least such distribution under stochastic dominant (partial) ordering. In previous works [18–20, 24], such an estimate is then used in place of the true (non-discretized) PLD for accounting and computing δ's (via Lemma 2.7 and Lemma 2.6).

6 This is referred to as *grid approximation* in [18–20].

A priori, it is not clear whether $\widetilde{\mathsf{PLD}}^{\uparrow}_{(A,B)}$ is even a valid PLD (for some pair of distributions). However, it not hard to prove that this is indeed the case:

Lemma 4.2. *There exists a pair $(P^{\uparrow}_{\mathrm{PB}}, Q^{\uparrow}_{\mathrm{PB}})$ of distributions such that $\widetilde{\mathsf{PLD}}^{\uparrow}_{(A,B)} = \mathsf{PLD}_{(P^{\uparrow}_{\mathrm{PB}}, Q^{\uparrow}_{\mathrm{PB}})}$.*

Proof. Let $P^{\uparrow}_{\mathrm{PB}}$ be defined by

$$P^{\uparrow}_{\mathrm{PB}}(\alpha_i) = \widetilde{\mathsf{PLD}}^{\uparrow}_{(A,B)}(\varepsilon_i)$$

for all $i \in \{0, \ldots, k\}$. It is clear that $P^{\uparrow}_{\mathrm{PB}}$ is a valid distribution.

Then, define $Q^{\uparrow}_{\mathrm{PB}}$ by

$$Q^{\uparrow}_{\mathrm{PB}}(\alpha) = P^{\uparrow}_{\mathrm{PB}}(\alpha)/\alpha,$$

for all $\alpha \in \mathcal{A} \setminus \{0\}$ and let $Q^{\uparrow}_{\mathrm{PB}}(0) = 1 - \sum_{\alpha \in \mathcal{A}\setminus\{0\}} Q^{\uparrow}_{\mathrm{PB}}(\alpha)$. To check that $Q^{\uparrow}_{\mathrm{PB}}$ is a valid distribution, it suffices to show that $Q^{\uparrow}_{\mathrm{PB}}(0) \geq 0$. This is true because

$$\begin{aligned}
\sum_{\alpha \in \mathcal{A}\setminus\{0\}} Q^{\uparrow}_{\mathrm{PB}}(\alpha) &= \sum_{i\in[k]} P^{\uparrow}_{\mathrm{PB}}(\alpha_i)/\alpha_i \\
&= \sum_{i\in[k]} \widetilde{\mathsf{PLD}}^{\uparrow}_{(A,B)}(\varepsilon_i)/\alpha_i \\
&= \sum_{i\in[k]} \mathsf{PLD}_{(A,B)}((\varepsilon_{i-1}, \varepsilon_i])/\alpha_i \\
&= \sum_{i\in[k]} \sum_{\substack{o\in\mathrm{supp}(B) \\ A(o)/B(o)\in(\alpha_{i-1},\alpha_i]}} A(o)/\alpha_i \\
&\leq \sum_{i\in[k]} \sum_{\substack{o\in\mathrm{supp}(B) \\ A(o)/B(o)\in(\alpha_{i-1},\alpha_i]}} B(o) \\
&\leq 1.
\end{aligned}$$

Finally, it is obvious from the definitions of $P^{\uparrow}_{\mathrm{PB}}, Q^{\uparrow}_{\mathrm{PB}}$ that $\mathsf{PLD}_{(P^{\uparrow}_{\mathrm{PB}}, Q^{\uparrow}_{\mathrm{PB}})} = \widetilde{\mathsf{PLD}}^{\uparrow}_{(A,B)}$. □

Combining the above lemma and the fact that $\widetilde{\mathsf{PLD}}^{\uparrow}_{(A,B)}$ dominates $\mathsf{PLD}_{(A,B)}$ with Lemma 4.1, we can conclude that our estimate is no worse than the PB estimate:

Corollary 4.3. *Let $(P^{\uparrow}, Q^{\uparrow})$ be as defined above. Then, for all $\alpha \geq 0$, we have $h_{(P^{\uparrow},Q^{\uparrow})}(\alpha) \leq D_{\alpha}(\widetilde{\mathsf{PLD}}^{\uparrow}_{(A,B)})$.*

Illustrations of the exact hockey-stick divergence and its pessimistic estimates from our approach and PB approach can be found in Figure 2. A more detailed evaluation of the error from the two approaches (after compositions) can be found in Section 6.

5 Optimistic PLDs with Finite Support

We next consider *optimistic* PLDs, i.e., $\mathsf{PLD}_{(P,Q)}$ dominated by a given $\mathsf{PLD}_{(A,B)}$. We start by showing that, unlike pessimistic PLDs for which there is the "best" possible choice (Lemma 4.1), there is no such a choice for optimistic PLDs:

Lemma 5.1. *There exists a pair (A, B) of distributions and a finite set $\mathcal{A}$ such that, for any pair $(P, Q) \preceq (A, B)$ such that $\mathrm{supp}(\mathsf{PLD}_{(P,Q)}) \subseteq \mathcal{E}$, there exists a pair $(P', Q') \preceq (A, B)$ such that $\mathsf{PLD}_{(P',Q')}$ is supported on $\mathcal{E}$ and $(P', Q') \not\preceq (P, Q)$.*

Proof. Let (A, B) be the result of the ε-DP binary randomized response, i.e.,

$$\begin{aligned}
A(0) = B(1) &= \frac{e^{\varepsilon}}{e^{\varepsilon}+1}, \\
A(1) = B(0) &= \frac{1}{e^{\varepsilon}+1}.
\end{aligned}$$

It is simple to verify that

$$h_{(A,B)}(\alpha) = \begin{cases} 1-\alpha & \text{if } \alpha \leq e^{-\varepsilon}, \\ \frac{e^{\varepsilon}}{e^{\varepsilon}+1} - \frac{\alpha}{e^{\varepsilon}+1} & \text{if } e^{-\varepsilon} < \alpha < e^{\varepsilon}, \\ 0 & \text{if } \alpha \geq \varepsilon. \end{cases}$$

Let $\mathcal{A}$ be $\{0, \alpha_1, \alpha_2, +\infty\}$ where $\alpha_1 = e^{-\varepsilon} - \gamma, \alpha_2 = e^{-\varepsilon} + \gamma$ for any $\gamma < \min\{e^{\varepsilon} - e^{-\varepsilon}, e^{-\varepsilon}\}$.

Let $h_1 : \mathcal{A} \cup \{+\infty\} \to [0, 1]$ be defined as

$$\begin{aligned}
h_1(0) &= 1, \\
h_1(\alpha_1) &= h_{(A,B)}(\alpha_1), \\
h_1(\alpha_2) &= 1 - \alpha_2, \\
h_1(+\infty) &= 0,
\end{aligned}$$

and let $\overline{h}_1$ be its piecewise-linear extension. It is again simple to verify that $\overline{h}_1$ satisfies the conditions in Lemma 3.2 and therefore $\overline{h}_1 = h_{(P_1,Q_1)}$ for some P_1, Q_1 such that $\mathsf{PLD}_{(P_1,Q_1)}$ is supported on $\mathcal{E}$. Furthermore, it can be checked from our definition that $(A, B) \succeq (P_1, Q_1)$.

Similarly, let $h_2 : \mathcal{A} \cup \{+\infty\} \to [0, 1]$ be defined as

$$\begin{aligned}
h_2(0) &= 1, \\
h_2(\alpha_1) &= 1 - e^{-\varepsilon} + \frac{\gamma}{e^{\varepsilon}+1}, \\
h_2(\alpha_2) &= h_{(A,B)}(\alpha_2), \\
h_2(+\infty) &= 0,
\end{aligned}$$

and let $\overline{h}_2$ be its piecewise-linear extension. Again, $\overline{h}_2 = h_{(P_2,Q_2)}$ for some P_2, Q_2 such that $\mathsf{PLD}_{(P_2,Q_2)}$ is supported on $\mathcal{E}$ and $(A,B) \succeq (P_2,Q_2)$.

Now, consider any $(P,Q) \preceq (A,B)$ such that $\mathsf{PLD}_{(P,Q)}$ is supported on $\mathcal{E}$. We claim that $(P,Q) \not\succeq (\hat{P}_1,\hat{Q}_1)$ or $(P,Q) \not\succeq (P_2,Q_2)$. To prove this, assume for the sake of contradiction that $(P,Q) \not\succeq (P_1,Q_1)$ and $(P,Q) \not\succeq (P_2,Q_2)$. This means that

$$h_{(P,Q)}(\alpha_1) \geq h_1(\alpha_1) = h_{(A,B)}(\alpha_1),$$
$$h_{(P,Q)}(\alpha_2) \geq h_2(\alpha_2) = h_{(A,B)}(\alpha_2).$$

From piecewise-linearity of $h_{(P,Q)}$ restricted to $[\alpha_1,\alpha_2]$ (Lemma 3.2), we then have $h_{(P,Q)}(e^\varepsilon) = \frac{1}{2}\left(h_{(P,Q)}(\alpha_1) + h_{(P,Q)}(\alpha_2)\right) > h_{(A,B)}(e^\varepsilon)$, a contradiction to the assumption that $(P,Q) \preceq (A,B)$. □

5.1 A Greedy and Convex Hull Construction

The previous lemma shows that, unfortunately, there is no canonical choice for an optimistic estimate for a given PLD. Due to this, we propose a simple greedy algorithm to construct an optimistic estimate for the PLD of a given pair (A,B). Similar to before, it will be more convenient to deal directly with the hockey-stick curve. Here we would like to construct $f : \mathcal{A} \to [0,1]$ such that its piecewise-linear extension $\overline{f}$ pointwise lower bounds $h_{(A,B)}$. The distribution P,Q (and $\mathsf{PLD}_{(P,Q)}$) can then be computed using Algorithm 1.

Our algorithm will assume that we can compute the derivative of h at any given point $\alpha \in \mathbb{R}_{\geq 0}$ (denoted by $h'(\alpha)$). We remark that, for many widely used mechanisms including Laplace and Gaussian mechanisms, the closed-formed formula for $h'(\alpha)$ can be easily computed.

First Greedy Attempt. Before describing our algorithm, let us describe an approach that does *not* work; this will demonstrate the hurdles we have to overcome. Consider the following simple greedy algorithm: start with $f(\alpha_0) = 0$ and, if we are currently at $f(\alpha_i)$, then find the largest possible $f(\alpha_{i+1})$ such that the line $f(\alpha_i), f(\alpha_{i+1})$ is below $h_{(A,B)}$. (In other words, the line $f(\alpha_i), f(\alpha_{i+1})$ is tangent to $h_{(A,B)}$.)

While this is a natural approach, there are two issues with this algorithm:

- First and more importantly, it is possible that at some discretization point $f(\alpha_{i+1})$ becomes negative! Obviously this invalidates the construction as f will not correspond to a hockey-stick curve.
- Secondly, each computation of tangent line requires several (and sequential) computations of the derivative h'—rendering the algorithm inefficient—and is also subject to possible numerical instability.

We remark that if we instead start from right (i.e., $f(\alpha_k)$) and proceed greedily to the left (in decreasing order of i), then the first issue will become that $f(\alpha_i)$ can be smaller than $[1-\alpha_i]_+$, which also makes it an invalid hockey-stick curve due to Lemma 2.1(iii). As will be explained below, we will combine these two directions of greedy together with a convex hull algorithm in our revised approach.

An Additional Assumption. For our algorithm, we will also need a couple of assumptions. The first one is that 1 belongs to the discretization set:

Assumption 5.2. $1 \in \mathcal{A}$ (or equivalently $0 \in \mathcal{E}$).

For the remainder of this section, we will use $i^* \in [k]$ to denote the index for which $\alpha_{i^*} = 1$ (i.e., $\varepsilon_{i^*} = 0$).

We show below that this assumption is necessary. When it does not hold, then it may simply be impossible to find an optimistic estimate at all:

Lemma 5.3. *There exists a pair (A,B) of distributions such that any $(P,Q) \preceq (A,B)$ satisfies $0 \in \mathrm{supp}(\mathsf{PLD}_{(P,Q)})$.*

Proof. Let $A = B$. We simply have $h_{(A,B)}(\alpha) = [1-\alpha]_+$. Due to Lemma 2.1(iii), we must have $h_{(P,Q)}(\alpha) = [1-\alpha]_+$. This simply implies $\mathsf{PLD}_{(P,Q)}(0) = 1$ as desired. □

Our Greedy + Convex Hull Algorithm. We are now ready to describe our final algorithm. The main idea is to not attempt to create the curve in one left-to-right or right-to-left sweep, but rather to simply generate a "candidate set" F_i for each $f(\alpha_i)$. (Such a set will in fact be a singleton for all i except $i = i^*$, for which $|F_i| \leq 2$, but we will refer to F_i's as sets here for simplicity of notation.) We then compute the convex hull of these points $\{(\alpha_i, f_i)\}_{i\in[k-1], f_i \in F_i}$ and take it (or more precisely its lower curve) as our optimistic estimate. This last step immediately ensures the convexity of our curve, which is required for it to be a valid hockey-stick curve (Lemma 2.1(i)).

To construct the candidate set F_i, we combine the left-to-right and right-to-left greedy approaches. Specifically, for $i = \{0,\dots,i^*-1\}$, we draw the tangent line of the true curve h at $h(\alpha_i)$ and let its intersection with the vertical line $\alpha = \alpha_{i+1}$ be $(\alpha_{i+1}, f^{\rightarrow}_{i+1})$; then we add

$f^{\rightarrow}_{i+1}$ into F_{i+1}. Similarly, for $i = \{k-1, \dots, i^* + 1\}$, we draw the tangent line of the true curve h at $h(\alpha_i)$ and let its intersection with the vertical line $\alpha = \alpha_{i-1}$ be $(\alpha_{i-1}, f^{\leftarrow}_{i-1})$; then we add $f^{\leftarrow}_{i-1}$ into F_{i-1}. At the very end points $i = 0$ and $i = k-1$, we also add 1 and 0 respectively to F_i.

Notice that this algorithm, unlike the previous (failed) greedy approach, only requires a calculation of h' at each $\alpha_i \in \mathcal{A} \setminus \{1, +\infty\}$, which can be done in parallel. Furthermore, efficient algorithms for convex hull are well known in the literature and can be used directly.

The complete and more precise description of our algorithm is given in Algorithm 2. We also note here that $|F_i| = 1$ for all $i \neq i^*$ and $|F_{i^*}| \leq 2$ (as the point constructed from the left may be different from the point from the right). Nonetheless, we write F_i's as sets for simplicity of notation. An illustration of the algorithm can be found in Figure 4.

Algorithm 2 Optimistic PLD Construction.

procedure OPTIMISTICPLD($h, \mathcal{A}$)
  for $i = 0, \dots i^* - 1$ **do**
    $f^{\rightarrow}_{i+1} = h(\alpha_i) + (\alpha_{i+1} - \alpha_i) \cdot h'(\alpha_i)$
  $f^{\rightarrow}_0 \leftarrow 1$ ▷ $h(\alpha_0) = 1$
  for $i = k-1, \dots i^* + 1$ **do**
    $f^{\leftarrow}_{i-1} = h(\alpha_i) - (\alpha_i - \alpha_{i-1}) \cdot h'(\alpha_i)$
  $f^{\leftarrow}_{k-1} \leftarrow 0$
  $H \leftarrow \text{ConvexHull}(\{(\alpha_i, f^{\rightarrow}_i)\}_{i \in \{0,\dots,i^*\}} \cup \{(\alpha_i, f^{\leftarrow}_i)\}_{i \in \{i^*,\dots,k-1\}})$
  for $i = 0, \dots, k-1$ **do**
    $(\alpha_i, f(\alpha_i)) \leftarrow$ lowest intersection point between H and the vertical line $\alpha = \alpha_i$
  $f(\alpha_k) \leftarrow 0$ ▷ $\alpha_k = +\infty$
  return DISCRETIZEPLD($f(\alpha_0), \dots, f(\alpha_k)$)

Having described our algorithm, we will now proof its correctness, i.e., that it outputs a pair of distributions dominated by the input pair.

Theorem 5.4. *Let $(P^{\downarrow}, Q^{\downarrow})$ denote the output of* OptimisticPLD($h, \mathcal{A}$) *where $h = h_{(A,B)}$. Then, under Assumption 5.2, we have $(P^{\downarrow}, Q^{\downarrow}) \preceq (A, B)$.*

To prove Theorem 5.4, it will be crucial to have the following lower bounds on the candidate points.

Lemma 5.5. *(i) For all $i \in \{0, \dots, i^*\}$, $f^{\rightarrow}(\alpha_i) \geq 1 - \alpha_i$.*
(ii) For all $i \in \{i^*, \dots, k-1\}$, $f^{\leftarrow}(\alpha_i) \geq 0$.
(iii) For all $i \in \{0, \dots, i^*\}$, $f^{\rightarrow}(\alpha_i) \leq h(\alpha_i)$.
(iv) For all $i \in \{i^*, \dots, k-1\}$, $f^{\leftarrow}(\alpha_i) \leq h(\alpha_i)$.

Proof. (i) The statement obviously holds for $i = 0$. Next, consider $i \in [i^*]$. From (ii) and (iii) of Lemma 2.1, we have $h'(0) \geq -1$. Furthermore, the convexity of h (Lemma 2.1(i)) implies that $h'(\alpha_{k-1}) \geq h'(0) \geq -1$. Therefore, we have

$$\begin{aligned} f^{\rightarrow}(\alpha_i) &= h(\alpha_{i-1}) + (\alpha_i - \alpha_{i-1}) \cdot h'(\alpha_{i-1}) \\ &\geq h(\alpha_{i-1}) - (\alpha_i - \alpha_{i-1}) \\ &\geq (1 - \alpha_{i-1}) - (\alpha_i - \alpha_{i-1}) \\ &= 1 - \alpha_i, \end{aligned}$$

where the second inequality is due to Lemma 2.1(iii).

(ii) The statement obviously holds for $i = k-1$. Next, consider $i \in \{i^*, \dots, k-2\}$. Since h is non-increasing (from Lemma 2.1(i)), we have $h'(\alpha_{i+1}) \leq 0$. Therefore, $f^{\leftarrow}(\alpha_i) \geq h(\alpha_{i+1}) \geq 0$.

(iii) The statement obviously holds for $i = 0$. For $i \in [i^*]$, the convexity of h (Lemma 2.1(i)) immediately implies that

$$f^{\rightarrow}(\alpha_i) = h(\alpha_{i-1}) + (\alpha_i - \alpha_{i-1}) \cdot h'(\alpha_{i-1}) \leq h(\alpha_i).$$

(iv) The statement obviously holds for $i = k-1$. For $i \in \{i^*, \dots, k-2\}$, the convexity of h (Lemma 2.1(i)) immediately implies that

$$\begin{aligned} f^{\rightarrow}(\alpha_i) &= h(\alpha_{i+1}) + (\alpha_{i+1} - \alpha_i) \cdot h'(\alpha_{i+1}) \\ &\leq h(\alpha_i). \end{aligned}$$

□

We are now ready to prove Theorem 5.4.

Proof of Theorem 5.4. We start by observing that the vertices of the convex hull H consists of the points $(\alpha_{\ell_0}, f^{\rightarrow}(\alpha_{\ell_0})), \dots, (\alpha_{\ell_q}, f^{\rightarrow}(\alpha_{\ell_q})), (\alpha_{\ell_{q+1}}, f^{\leftarrow}(\alpha_{\ell_{q+1}})), \dots, (\alpha_{\ell_m}, f^{\leftarrow}(\alpha_{\ell_m}))$, where $0 = \ell_0 < \cdots < \ell_m = k-1$ and $\ell_q \leq i^* \leq \ell_{q+1}$.

First, we have to show that $f(\alpha_0), \dots, f(\alpha_k)$ constitute a valid input to the DISCRETIZEPLD algorithm. Per Lemma 3.2, we only need to show that (i) $\overline{f}$ is non-increasing, (ii) $\overline{f}$ is convex, and (iii) $\overline{f}(\alpha) \geq [1 - \alpha]_+$ for all $\alpha \in \mathbb{R}_{\geq 0}$. Notice also that $\overline{f}$ is simply the piecewise-linear curve connecting $(\alpha_{\ell_0}, f^{\rightarrow}(\alpha_{\ell_0})), \dots, (\alpha_{\ell_q}, f^{\rightarrow}(\alpha_{\ell_q})), (\alpha_{\ell_{q+1}}, f^{\leftarrow}(\alpha_{\ell_{q+1}})), \dots, (\alpha_{\ell_m}, f^{\leftarrow}(\alpha_{\ell_m}))$.

To see that (i) holds, observe that the second-rightmost point in the convex hull must be (α_j, f_j) for some $j \in \{0, \dots, k-2\}$ where $f_j = f^{\rightarrow}_j$ or $f_j = f^{\leftarrow}_j$. In either case, Lemma 5.5 implies that $f_j \geq 0 = f^{\leftarrow}(\alpha_{k-1})$. Since $\overline{f}$ is the lower curve of the convex hull H and

$(\alpha_j, f_j), (\alpha_{k-1}, 0)$ is its rightmost segment, we can conclude that $\overline{f}$ is non-increasing in the range $[\alpha_0, \alpha_{k-1}]$. Finally, since we simply have $\overline{f}(\alpha) = 0$ for all $\alpha > \alpha_{k-1}$, it is also non-increasing in the range $[\alpha_{k-1}, +\infty)$, thereby proving (i).

As for (ii), since $\overline{f}|_{[\alpha_0, \alpha_{k-1}]}$ forms the lower boundary of the convex hull H, $\overline{f}$ is convex in the range $[\alpha_0, \alpha_{k-1}]$. Again, since we simply have $\overline{f}(\alpha) = 0$ for all $\alpha > \alpha_{k-1}$, we can conclude that it is convex for the entire range $[0, +\infty)$.

Finally, for (iii), Lemma 5.5 states that all the points $\{(\alpha_i, f_i^{\rightarrow})\}_{i \in \{0,\dots,i^*\}} \cup \{(\alpha_i, f_i^{\leftarrow})\}_{i \in \{i^*,\dots,k-1\}}$ is above the curve $\alpha \mapsto [1-\alpha]_+$. Since $\overline{f}$ is in the convex hull H, we can conclude that $\overline{f}$ also lies above this curve, as desired.

Now that we have proved that $f(\alpha_0), \dots, f(\alpha_k)$ is a valid input to the DiscretizePLD algorithm (and therefore the output $(P^{\uparrow}, Q^{\uparrow})$ is a pair of valid probability distributions), we will next show that $(P^{\uparrow}, Q^{\uparrow}) \preceq (A, B)$. This is equivalent to showing that $\overline{f}(\alpha) \le h(\alpha)$ for all $\alpha \ge \mathbb{R}_{\ge 0} \cup \{+\infty\}$. To prove this, we consider three cases based on the value of α. For brevity, we say that a curve C_1 is *below* a curve C_2 when they share the same domain Ω and $C_1(o) \le C_2(o)$ for all $o \in \Omega$.

- Case I: $\alpha \ge \alpha_{k-1}$. In this case, $\overline{f}(\alpha) = 0 \le h(\alpha)$.
- Case II: $\alpha \in [1, \alpha_{k-1})$. Suppose that $\alpha \in [\alpha_i, \alpha_{i+1})$. Let L_1 denote the line segment from $(\alpha_i, f^{\leftarrow}(\alpha_i))$ to $(\alpha_{i+1}, f^{\leftarrow}(\alpha_{i+1}))$, and L_2 denote the line segment from $(\alpha_i, f^{\leftarrow}(\alpha_i))$ to $(\alpha_{i+1}, h(\alpha_{i+1}))$. From Lemma 5.5(iv), L_1 is below L_2. Furthermore, since $\overline{f}|_{[\alpha_i, \alpha_{i+1})}$ is a lower boundary of the convex hull H containing L_1, it must also be below L_1. Therefore, we have

$$\begin{aligned}\overline{f}(\alpha) \le L_2(\alpha) &\le L_1(\alpha) \\ &= h(\alpha_{i+1}) - (\alpha_{i+1} - \alpha) h'(\alpha_{i+1}) \\ &\le h(\alpha),\end{aligned}$$

  where the last inequality follows from convexity of h.
- Case III: $\alpha \in [0, 1)$. Suppose that $\alpha \in [\alpha_i, \alpha_{i+1})$. Similarly to the previous case, let L_3 denote the line segment from $(\alpha_i, f^{\rightarrow}(\alpha_i))$ to $(\alpha_{i+1}, f^{\rightarrow}(\alpha_{i+1}))$, and L_4 denote the line segment from $(\alpha_i, h(\alpha_i))$ to $(\alpha_{i+1}, f^{\rightarrow}(\alpha_{i+1}))$. From Lemma 5.5(iii), L_3 is below L_4. Furthermore, since $\overline{f}|_{[\alpha_i, \alpha_{i+1})}$ is a lower boundary of the convex hull H containing L_3, it must also be below L_3. Therefore, we have

$$\begin{aligned}\overline{f}(\alpha) \le L_3(\alpha) &\le L_4(\alpha) \\ &= h(\alpha_i) + (\alpha_{i+1} - \alpha) h'(\alpha_i) \\ &\le h(\alpha),\end{aligned}$$

  where the last inequality follows from convexity of h.

As a result, we can conclude that $(P^{\uparrow}, Q^{\uparrow}) \preceq (A, B)$, completing our proof. □

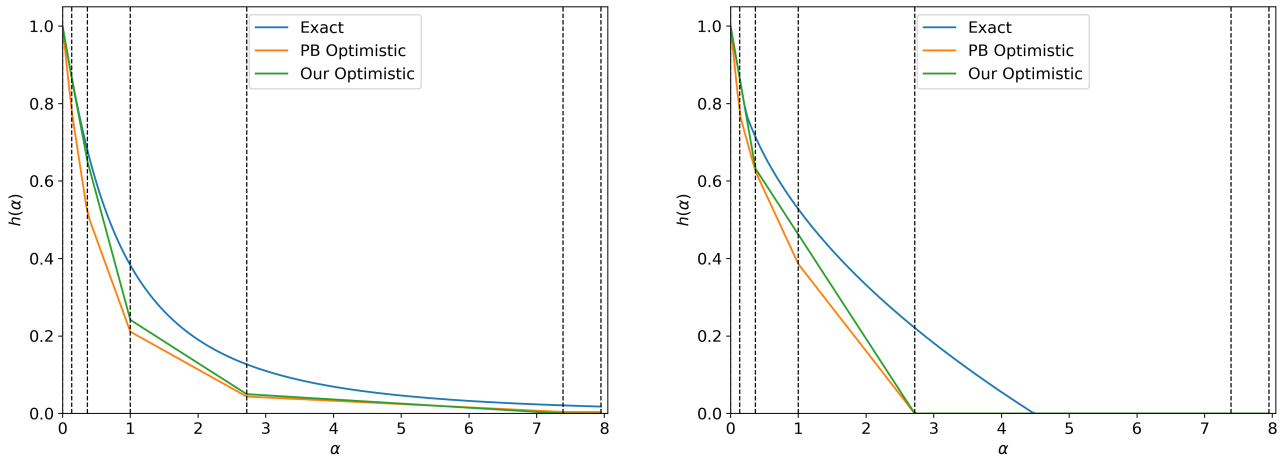
5.2 Comparison to Privacy Loss Buckets

Similar to Section 4.1, PB [24] can also be applied for optimistic-estimate: let $\widetilde{\mathsf{PLD}}^{\downarrow}_{(A,B)}$ be the probability distribution where

$$\widetilde{\mathsf{PLD}}^{\downarrow}_{(A,B)}(\varepsilon_{i-1}) = \mathsf{PLD}_{(A,B)}([\varepsilon_{i-1}, \varepsilon_i)),$$

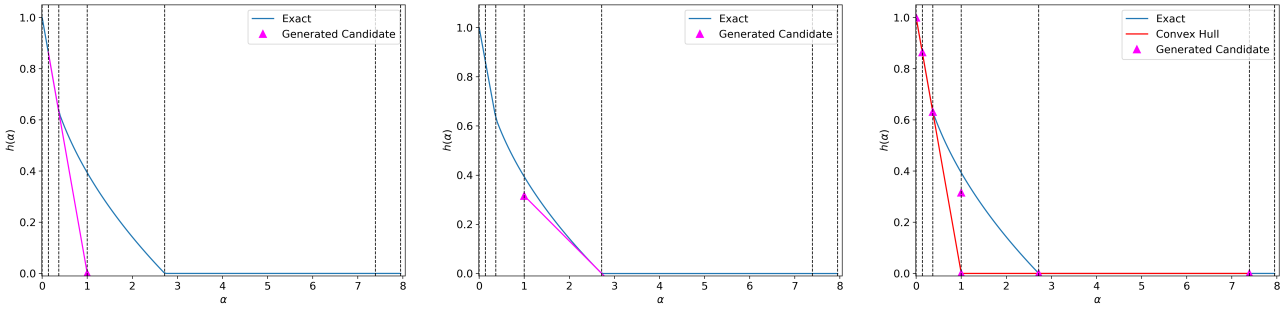
for all $i \in [k]$. That is, $\widetilde{\mathsf{PLD}}^{\downarrow}_{(A,B)}$ is a probability distribution on $\mathcal{E}$ that is stochastically dominated by $\mathsf{PLD}_{(A,B)}$; furthermore, $\widetilde{\mathsf{PLD}}^{\downarrow}_{(A,B)}$ is the greatest such distribution under stochastic dominant (partial) ordering. It is important to note that, unlike the pessimistic-PB estimate, the optimistic-PB estimate $\widetilde{\mathsf{PLD}}^{\downarrow}_{(A,B)}$ is *not* necessarily a valid PLD for some pair of distributions. This can easily be seen by, e.g., taking a PLD of any ε-DP mechanism for finite ε and let $\mathcal{E} = \{-\infty, +\infty\}$; the optimistic-PB estimate puts all of its mass at 0, which is clearly not a valid PLD.

We present illustrations of our optimistic PLD and optimistic-PB estimates in Figure 3. Recall that in the pessimistic case, we can show that the pessimistic-PB estimate is no better than our approach (Corollary 4.3). Although we observe similar behaviours in the optimistic case in simple examples (e.g. Figure 3) and also in our experiments in Section 6, this unfortunately does not hold in general. Indeed, if $\mathsf{PLD}_{(A,B)}$ has a non-zero mass at $+\infty$ (or equivalently $h(+\infty) \neq 0$), then the optimistic-PB estimate still keeps this mass while our does not. The latter is because we set $f(+\infty) = 0$ in Algorithm 2. Note here that we cannot set $f(+\infty) = h(+\infty)$ here because the monotonicity may not hold anymore; it is possible that $f^{\rightarrow}(\alpha_i) < h(+\infty)$ for some $i \in [i^*]$. Such examples highlight the challenge in finding a good optimistic estimate (especially in light of the non-existence of the best one, i.e., Lemma 5.1), and we provide further discussion regarding this in Section 7.



(a) Gaussian Mechanism **(b)** Laplace Mechanism

Fig. 3. Illustrations of the hockey-stick curves of the Gaussian and Laplace mechanisms , and their optimistic estimates from our approach and the Privacy Bucket (PB) approach of [24]. The setting of parameters and labels are similar to Figure 2.



(a) One Step of Candidate Generation for $i < i^*$ **(b)** One Step of Candidate Generation for $i > i^*$ **(c)** All Candidate Points and Its Convex Hull

Fig. 4. Illustrations of our optimistic PLD construction algorithm (Algorithm 2) for Laplace mechanism with noise multiplier 1. Figure 4a demonstrates one step of how the candidate points are generated when $i < i^*$. Specifically, a line tangent to the hockey-stick curve is drawn at each point $(\alpha_i, h(\alpha_i))$; the intersections with the vertical line at α_{i+1} give the candidate points $(\alpha_{i+1}, f^{\rightarrow}_{i+1})$. Similarly, Figure 4b shows such a step for $i > i^*$; in this case, the same line is drawn and its intersection with the vertical line at α_{i-1} give the candidate points $(\alpha_{i-1}, f^{\leftarrow}_{i-1})$. Figure 4c shows all the candidate points generated together with its convex hull, which we use as our optimistic estimate.

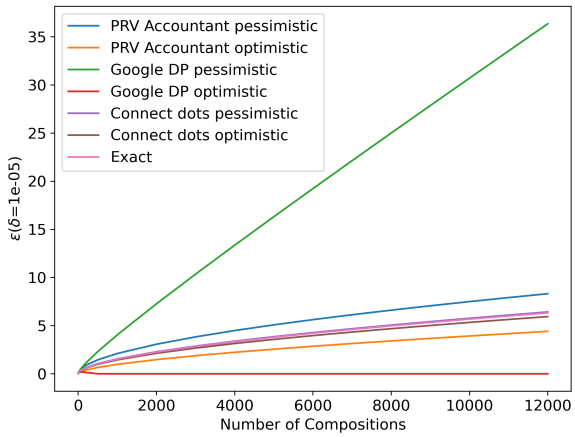


Fig. 5. ε vs Number of Compositions of the Gaussian mechanism with noise scale 80. All methods evaluated with the same discretization interval of 0.005.

6 Evaluation

We compare our algorithm with the Privacy Buckets algorithm [24] as implemented in the Google DP library[7], and the algorithm of Gopi et al. [15] implemented in Microsoft PRV Accountant.[8] Gopi et al.'s algorithm does not fit into the pessimistic/optimistic framework as described in Section 2.5. Instead, their algorithm uses an approximation of PLD that is neither optimistic nor pessimistic and uses a concentration bound to derive pessimistic and optimistic estimates. Indeed, their approximate distribution maintains the same expectation as the true PLD, which is the main ingredient in their improvement over previous work.

As a first cut, we evaluate pessimistic and optimistic estimates on the privacy parameter ε, for a fixed value of $\delta = 10^{-5}$, for varying number of compositions of the Gaussian mechanism with noise scale 80, while comparing our approach to the two other implementations mentioned above. We use the same discretization interval to evaluate each algorithm. The reason for choosing the Gaussian mechanism is that the exact value of ε can be computed explicitly. We find that the estimates given by our approach are the tightest.

Remark 6.1. For any specified discretization interval, each algorithm has a different choice of how many discretization points are included in $\mathcal{E}$. Our implementation uses the same set of discretization points as used by the Google DP implementation. The number of discretization points increases with the number of self compositions (we use the Google DP implementation to perform self composition[9]). On the other hand, Microsoft PRV Accountant chooses a number of discretization points, depending on the number of compositions desired, and this number does not change after self composition. In all the evaluation experiments mentioned in this paper, we find that the number of discretization points in our approach are lower than the number of discretization points in the PRV Accountant (even after composition).

Our main evaluation involves computing pessimistic and optimistic estimates on the privacy parameter ε, for a fixed value of δ, for varying number of compositions of the Poisson sub-sampled Gaussian mechanism and comparing our approach to each of the two other implementations. Note that this particular mechanism is quite popular in that it captures the privacy analysis of DP-SGD where the number of compositions is equal to the number of iterations of the training algorithm, and the subsampling rate is equal to the fraction of the batch size divided by the total number of training examples [1]. In particular, we consider the Gaussian mechanism with noise scale 1, Poisson-subsampled with probability 0.01. We compare against each competing algorithm twice, once where both algorithms use the same discretization interval, and once where our approach uses a larger discretization interval than the competing algorithm. We additionally plot the running time required for this computation for each number of compositions; we ran the evaluation for each number of compositions 20× and plot the mean running time along with a shaded region indicating 25th–75th percentiles of running time.

7 Even though there are several other papers [18–20] that build on PB, all of them still use the same PB-based approximation, with the differences being how the truncation is computed for FFT. We use the implementation in the Google DP library github.com/google/differential-privacy/tree/main/python

8 Implementation from github.com/microsoft/prv_accountant

9 We found that the Google DP implementation has a significantly worse running time when computing optimistic estimates, due to lack of truncation. We modify the self-composition method in the Google DP library to incorporate truncation when computing optimistic estimates, and evaluate both ours and Google DP implementation with this minor modification. These do not change the estimates significantly, but drastically reduce the running time.

Comparison with Google DP.
The comparison with Google DP is presented in Figure 6. Figures 6a and 6c compare the ε's and runtimes for both methods using the same discretization interval, and finds that our method gives a significantly tighter estimate for a mildly larger running time. Figures 6b and 6d compare the ε's and runtimes for both methods with different discretization intervals, and using a discretization interval that is $66.66\times$ larger, our method gives comparable estimates, with a drastic speed-up ($\sim 300\times$).

Comparison with Microsoft PRV Accountant.
The comparison with Microsoft PRV Accountant is presented in Figure 7. Figures 7a and 7c compare the ε's and runtimes for both methods using the same discretization interval, and finds that our method gives a significantly tighter estimate with already shorter running time. Figures 7b and 7d compare the ε's and runtimes for both methods with different discretization intervals, and using a discretization interval that is $6.66\times$ larger, our method gives comparable estimates, with an even larger speed-up.

In Appendix B, we perform a similar evaluation for the Poisson-subsampled Laplace mechanism.

7 Discussion & Open Problems

In this work, we have proposed a novel approach to pessimistic and optimistic estimates for PLDs, which outperforms previous approaches under similar discretization intervals, and allows for a more compact representation of PLDs while retaining similar error guarantees. There are still several interesting future directions that one could consider.

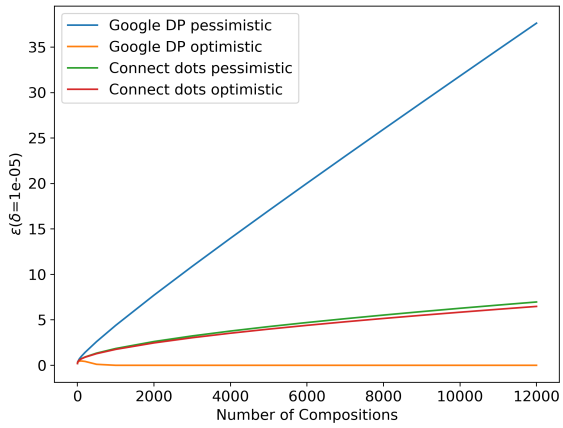
As we have proved in Lemma 5.1, there is no unique "best" way to pick an optimistic estimate, and we proposed a greedy algorithm (Algorithm 2) for this task. However, it is difficult to determine how good this greedy algorithm is in general. Instead, it might also be interesting to find $(P^{\downarrow}, Q^{\downarrow}) \preceq (A, B)$ that minimizes a certain objective involving $h_{(P^{\downarrow},Q^{\downarrow})}$ and $h_{(A,B)}$. For example, one could consider the area between the two curves, or the Fréchet distance between them. An intriguing direction here is to determine (i) which objective captures the notion of "good approximation" better in terms of composition, and (ii) for a given objective, whether there is an efficient algorithm to compute such $(P^{\downarrow}, Q^{\downarrow})$. We remark that for some objectives, such as the area between the two curves, it is possible to discretize the candidate values for each $f(\alpha_i)$ and use dynamic programming in an increasing order of i (with the state being $f(\alpha_{i-1}), f(\alpha_i)$). Even for these objectives, it remains interesting to determine whether such discretization is necessary and whether more efficient algorithms exist.

Also related is the question of how to theoretically explain our experimental findings (Section 6). Although we see significant numerical improvements, it is intriguing to understand theoretically where these improvements come from and which properties of PLDs govern how big such improvements are. More broadly, given an estimate of a PLD, how can we quantify how "good" it is? Previous work (e.g., [15, 19]) has obtained certain theoretical bounds on the errors; it would be interesting to investigate whether these bounds can help answer the aforementioned question.

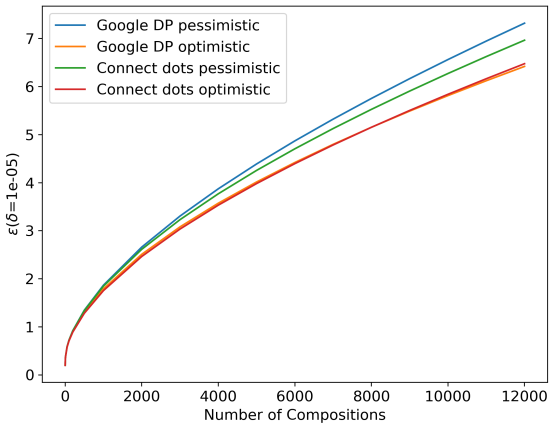
Furthermore, the entire line of work on PLD-based accounting [18–20, 24], including this paper, has so far considered only *non-interactive* compositions, meaning that the mechanisms that are run in subsequent steps cannot be changed based on the outputs from the previous steps. This is not a coincidence: interactive composition is highly non-trivial and in fact it is known that the advanced composition theorem (which is even a more specific form of PLD) does not hold in this regime [28]. Several solutions have been proposed here, such as modified formulae for advanced compositions [28, 31] and Renyi DP [14, 21]. However, as discussed earlier, these methods may be loose even in the non-interactive setting, which is the original motivation for PLD-based accounting. Therefore, it would be interesting to understand whether there is a tighter method similar to PLDs that also works in the interactive setting.

Acknowledgments

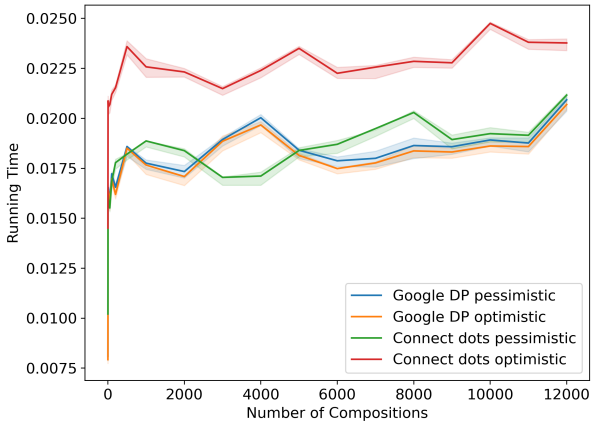
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.



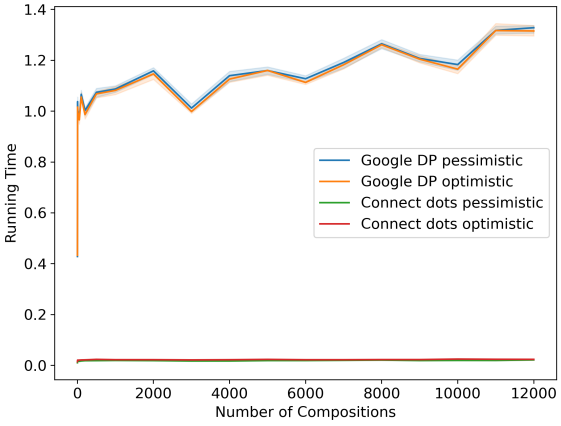
(a) ε comparison with discretization interval 0.005 for both this work and Google DP



(b) ε comparison with discretization interval 0.005 for this work, 0.000075 for Google DP

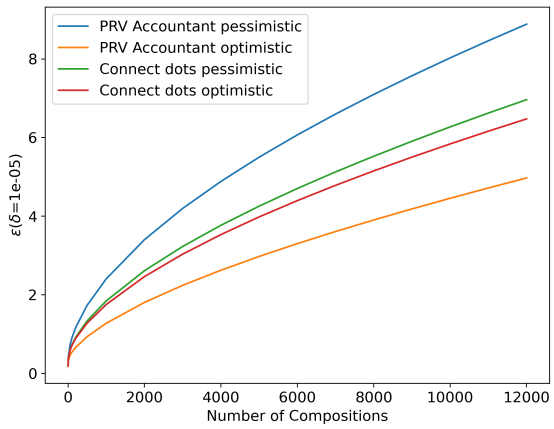


(c) Running time comparison with discretization interval 0.005 for both this work and Google DP (shaded region indicates 25-75 percentile range across 20 independent runs)

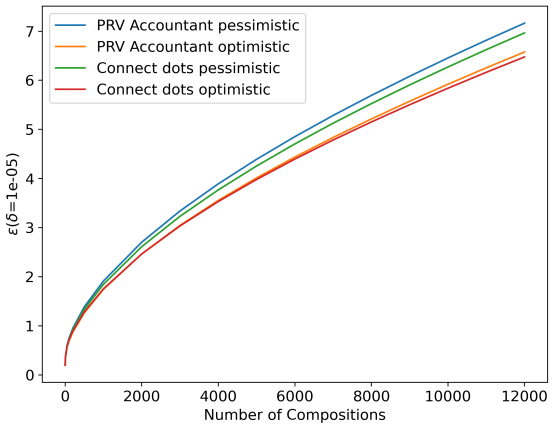


(d) Running time comparison with discretization interval 0.005 for this work, 0.000075 for Google DP (shaded region indicates 25-75 percentile range across 20 independent runs)

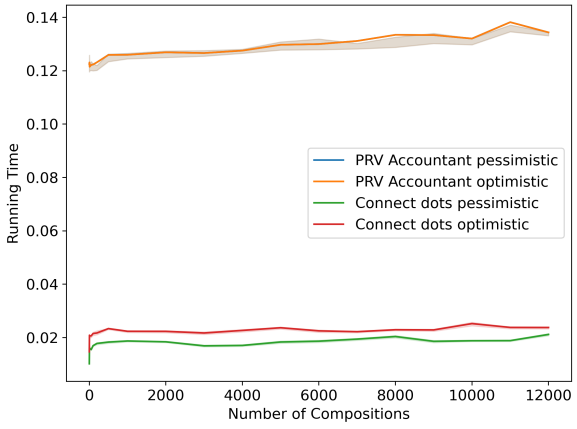
Fig. 6. Computation of pessimistic/optimistic estimates of the privacy parameter ε (for fixed parameter $\delta = 10^{-5}$) for self-composition of the Gaussian mechanism with noise scale 1, Poisson-subsampled with probability 0.01, using the Google DP implementation of the PB approach [24] vs. our approach. Figures 6a and 6c compare the ε's and runtimes for both methods using the same discretization interval. Figures 6b and 6d compare the ε's and runtimes for both methods with different discretization intervals.



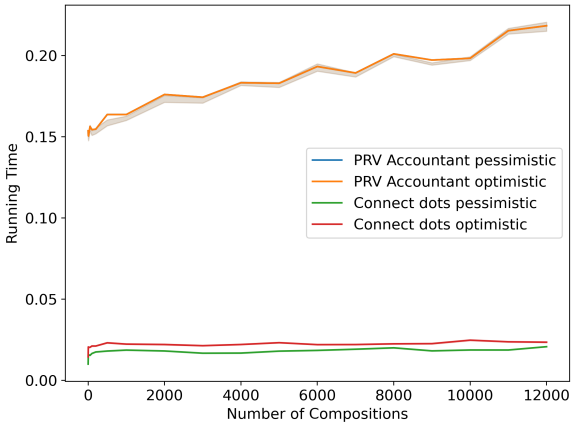
(a) ε comparison with discretization interval 0.005 for both this work and PRV Accountant



(b) ε comparison with discretization interval 0.005 for this work, 0.00075 for PRV Accountant (shaded region indicates 25-75 percentile range across 20 independent runs)



(c) Running time comparison with discretization interval 0.005 for both this work and PRV Accountant (shaded region indicates 25-75 percentile range across 20 independent runs)



(d) Running time comparison with discretization interval 0.005 for this work, 0.00075 for PRV Accountant

Fig. 7. Computation of pessimistic/optimistic estimates of the privacy parameter ε (for fixed parameter $\delta = 10^{-5}$) for self-composition of the Gaussian mechanism with noise scale 1, Poisson-subsampled with probability 0.01, using the Microsoft PRV Accountant [15] vs. our approach. We note that the pessimistic and optimistic curves of the Microsoft PRV Accountant [15] are identical in Figure 7c and Figure 7d.

References

[1] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *CCS*, pages 308–318, 2016.

[2] John M Abowd. The US Census Bureau adopts differential privacy. In *KDD*, pages 2867–2867, 2018.

[3] Apple Differential Privacy Team. Learning with privacy at scale. *Apple Machine Learning Journal*, 2017.

[4] Zhiqi Bu, Sivakanth Gopi, Janardhan Kulkarni, Yin Tat Lee, Judy Hanwen Shen, and Uthaipon Tantipongpipat. Fast and memory efficient differentially private-SGD via JL projections. In *NeurIPS*, 2021.

[5] Mark Bun, Cynthia Dwork, Guy N Rothblum, and Thomas Steinke. Composable and versatile privacy via truncated CDP. In *STOC*, pages 74–86, 2018.

[6] Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *TCC*, pages 635–658, 2016.

[7] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting telemetry data privately. In *NeurIPS*, pages 3571–3580, 2017.

[8] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *EUROCRYPT*, pages 486–503, 2006.

[9] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, pages 265–284, 2006.

[10] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.

[11] Cynthia Dwork and Guy N Rothblum. Concentrated differential privacy. *arXiv:1603.01887*, 2016.

[12] Cynthia Dwork, Guy N Rothblum, and Salil Vadhan. Boosting and differential privacy. In *FOCS*, pages 51–60, 2010.

[13] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *CCS*, pages 1054–1067, 2014.

[14] Vitaly Feldman and Tijana Zrnic. Individual privacy accounting via a Rényi filter. In *NeurIPS*, 2021.

[15] Sivakanth Gopi, Yin Tat Lee, and Lukas Wutschitz. Numerical composition of differential privacy. In *NeurIPS*, 2021.

[16] Andy Greenberg. Apple's "differential privacy" is about collecting your data – but not your data. *Wired, June*, 13, 2016.

[17] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The composition theorem for differential privacy. In *ICML*, pages 1376–1385, 2015.

[18] Antti Koskela and Antti Honkela. Computing differential privacy guarantees for heterogeneous compositions using FFT. *arXiv:2102.12412*, 2021.

[19] Antti Koskela, Joonas Jälkö, and Antti Honkela. Computing tight differential privacy guarantees using FFT. In *AISTATS*, pages 2560–2569, 2020.

[20] Antti Koskela, Joonas Jälkö, Lukas Prediger, and Antti Honkela. Tight differential privacy for discrete-valued mechanisms and for the subsampled Gaussian mechanism using FFT. In *AISTATS*, pages 3358–3366, 2021.

[21] Mathias Lécuyer. Practical privacy filters and odometers with Rényi differential privacy and applications to differentially private deep learning. *arXiv:2103.01379*, 2021.

[22] Google's Differential Privacy Libraries. DP Accounting Library. https://github.com/google/differential-privacy/tree/main/python/dp_accounting, 2020.

[23] Antti Koskela. Lukas Prediger. Code for computing tight guarantees for differential privacy. https://github.com/DPBayes/PLD-Accountant, 2020.

[24] Sebastian Meiser and Esfandiar Mohammadi. Tight on budget? Tight bounds for r-fold approximate differential privacy. In *CCS*, pages 247–264, 2018.

[25] Microsoft. A fast algorithm to optimally compose privacy guarantees of differentially private (DP) mechanisms to arbitrary accuracy. https://github.com/microsoft/prv_accountant, 2021.

[26] Ilya Mironov. Rényi differential privacy. In *CSF*, pages 263–275, 2017.

[27] Jack Murtagh and Salil Vadhan. The complexity of computing the optimal composition of differential privacy. In *TCC*, pages 157–175, 2016.

[28] Ryan M. Rogers, Salil P. Vadhan, Aaron Roth, and Jonathan R. Ullman. Privacy odometers and filters: Pay-as-you-go composition. In *NIPS*, pages 1921–1929, 2016.

[29] Stephen Shankland. How Google tricks itself to protect Chrome user privacy. *CNET, October*, 2014.

[30] David M Sommer, Sebastian Meiser, and Esfandiar Mohammadi. Privacy loss classes: The central limit theorem in differential privacy. *PoPETS*, 2019(2):245–269, 2019.

[31] Justin Whitehouse, Aaditya Ramdas, Ryan Rogers, and Steven Wu. Improved privacy filters and odometers: Time-uniform bounds in privacy composition. In *TPDP*, 2021.

[32] Yuqing Zhu, Jinshuo Dong, and Yu-Xiang Wang. Optimal accounting of differential privacy via characteristic function. In *AISTATS*, pages 4782–4817, 2022.

A Proofs

Lemma 2.6. *For any pair (P, Q) of discrete distributions and $\varepsilon \in \mathbb{R} \cup \{-\infty, +\infty\}$, we have*

$$D_{e^\varepsilon}(P||Q) := \sum_{\varepsilon' \in \mathrm{supp}(\mathsf{PLD}_{(P,Q)})} [1 - e^{\varepsilon - \varepsilon'}]_+ \cdot \mathsf{PLD}_{(P,Q)}(\varepsilon').$$

Proof. We have from the definition of hockey-stick divergence that

$$\begin{aligned} D_{e^\varepsilon}(P||Q) &= \sum_\omega [P(\omega) - e^\varepsilon \cdot Q(\omega)]_+ \\ &= \sum_\omega [1 - e^{\varepsilon - \log(P(\omega)/Q(\omega))}]_+ \cdot P(\omega) \\ &= \sum_{\varepsilon' \in \mathrm{supp}(\mathsf{PLD}_{(P,Q)})} [1 - e^{\varepsilon - \varepsilon'}]_+ \cdot \mathsf{PLD}_{(P,Q)}(\varepsilon') \end{aligned}$$

where the last line follows from the fact that

$$\mathsf{PLD}_{(P,Q)}(\varepsilon') := \sum_{\omega : \log(P(\omega)/Q(\omega)) = \varepsilon'} P(\omega)\,. \qquad \square$$

B Evaluation of Poisson-Subsampled Laplace Mechanism

Similar to Section 6, we compute pessimistic and optimistic estimates on the privacy parameter ε, for a fixed value of δ, for varying number of compositions of the Poisson sub-sampled Laplace mechanism and comparing our approach to the Google DP implementation.[10] In particular, we consider the Laplace mechanism with noise scale 5, Poisson-subsampled with probability 0.01. We compare against Google DP implementation twice, once where both algorithms use the same discretization interval, and once where our approach uses a larger discretization interval than the competing algorithm. We additionally plot the running time required for this computation for each number of compositions; we ran the evaluation for each number of compositions 20× and plot the mean running time along with a shaded region indicating 25th–75th percentiles of running time.

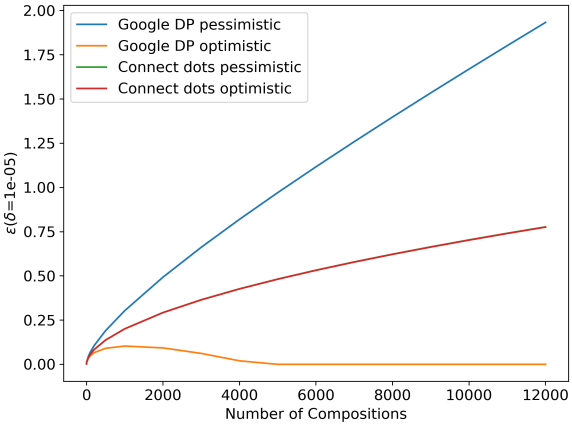
The comparison with Google DP is presented in Figure 8. Figures 6a and 8c compare the ε's and runtimes for both methods using the same discretization interval, and finds that our method gives a significantly tighter estimate for a mildly larger running time. Figures 8b and 8d compare the ε's and runtimes for both methods with different discretization intervals, and using a discretization interval that is 100× larger, our method gives comparable estimates, with a significant running time speed-up ($\sim 75\times$).

10 we were unable to compare against Microsoft PRV Accountant, since their implementation does not have support for the Laplace mechanism yet.

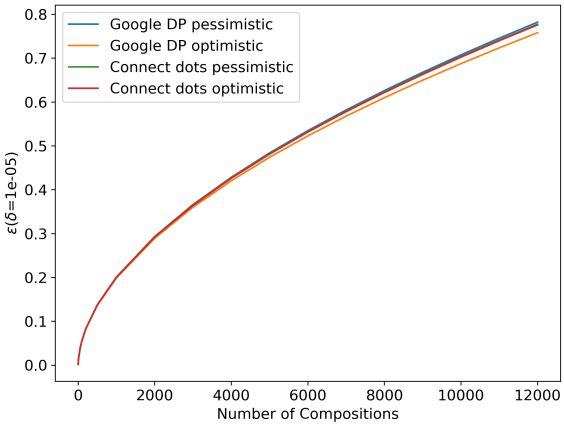
C Inaccuracies from Floating-Point Arithmetic

We briefly discuss the errors due to floating-point arithmetic. In our implementation, we use the default float datatype in python, which conforms to IEEE-754 "double precision". Roughly speaking, this means that the resolution of each floating-point number of $2^{-53} \approx 1.1 \cdot 10^{-16}$. The number of operations performed in our algorithms scales linearly with the support size of the (discretized) PLD, which is less than 10^4 in all of our experiments. Therefore, a rough heuristic suggests that the numerical error for δ here would be less than 10^{-11}. We stress however that this is just a heuristic and is not a formal guarantee: achieving a formal guarantee is much more complicated, e.g., our optimistic algorithm requires computing a convex hull and one would have to formalize how the numerical error from convex hull computation affects the final δ.

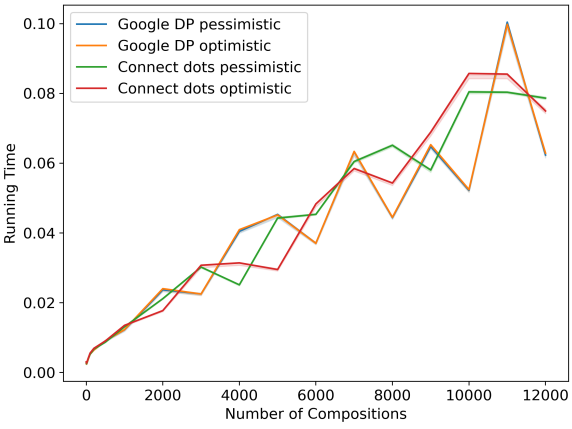
Finally, we also remark that, while Gopi et al. [15, Appendix A] note that they experience numerical issues around $\delta \approx 10^{-9}$, we do not experience the same issues in our algorithm even for similar setting of parameters even for δ as small as 10^{-12}.



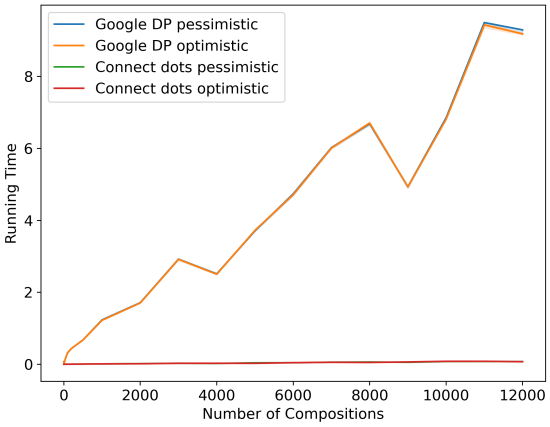
(a) ε comparison with discretization interval 0.0002 for both this work and Google DP



(b) ε comparison with discretization interval 0.0002 for this work, 0.000002 for Google DP (shaded region indicates 25-75 percentile range across 20 independent runs)



(c) Running time comparison with discretization interval 0.0002 for both this work and Google DP (shaded region indicates 25-75 percentile range across 20 independent runs)



(d) Running time comparison with discretization interval 0.0002 for this work, 0.000002 for Google DP

Fig. 8. Computation of pessimistic/optimistic estimates of the privacy parameter ε (for fixed parameter $\delta = 10^{-5}$) for self-composition of the Laplace mechanism with noise scale 1, Poisson-subsampled with probability 0.01, using the Google DP implementation of the PB approach [24] vs. our approach.Figures 8a and 8c compare the ε's and runtimes for both methods using the same discretization interval. Figures 8b and 8d compare the ε's and runtimes for both methods with different discretization intervals.