Composable and Versatile Privacy via Truncated CDP

Mark Bun
Princeton University
Princeton, NJ, USA
mbun@cs.princeton.edu

Cynthia Dwork
Harvard University
Cambridge, MA, USA
dwork@seas.harvard.edu

Guy N. Rothblum
Weizmann Institute of Science
Rehovot, Israel
rothblum@alum.mit.edu

Thomas Steinke
IBM Research – Almaden
San Jose, CA, USA
tcdp@thomas-steinke.net

ABSTRACT

We propose *truncated* concentrated differential privacy (tCDP), a refinement of differential privacy and of concentrated differential privacy. This new definition provides robust and efficient composition guarantees, supports powerful algorithmic techniques such as privacy amplification via sub-sampling, and enables more accurate statistical analyses. In particular, we show a central task for which the new definition enables exponential accuracy improvement.

CCS CONCEPTS

• **Theory of computation → Design and analysis of algorithms**;

KEYWORDS

differential privacy, algorithmic stability, subsampling

ACM Reference Format:
Mark Bun, Cynthia Dwork, Guy N. Rothblum, and Thomas Steinke. 2018. Composable and Versatile Privacy via Truncated CDP. In *Proceedings of 50th Annual ACM SIGACT Symposium on the Theory of Computing (STOC'18).* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3188745.3188946

1 INTRODUCTION

Differential privacy (DP) is a mathematically rigorous definition of privacy, which is tailored to analysis of large datasets and is equipped with a formal measure of privacy loss [9, 12]. Differentially private algorithms cap the permitted privacy loss in any execution of the algorithm at a pre-specified parameter ε, providing a concrete privacy/utility tradeoff. A signal strength of differential privacy is the ability to reason about cumulative privacy loss under *composition* of multiple analyses.

Roughly speaking, differential privacy ensures that the outcome of any analysis on a dataset x is distributed very similarly to the outcome on any neighboring dataset x' that differs from x in just one element (corresponding to one individual). That is, differentially private algorithms are randomized, and the *max divergence* between these two distributions (the maximum log-likelihood ratio for any event) is bounded by the privacy parameter ε. This absolute guarantee on the maximum privacy loss is now sometimes referred to as "pure" differential privacy.

A popular relaxation, "approximate" or (ε, δ)-differential privacy [10], roughly guarantees that with probability at least $1 - \delta$ the privacy loss does not exceed ε. This relaxation allows a δ probability of catastrophic privacy failure, and thus δ is typically taken to be "cryptographically" small. Although small values of δ come at a price in privacy, the relaxation nevertheless frequently permits asymptotically better accuracy than pure differential privacy (for the same value of ε). Indeed, the central advantage of relaxing the guarantee is that it permits an improved and asymptotically tight analysis of the cumulative privacy loss incurred by composition of multiple (pure or approximate) differentially private mechanisms [15].

Composition is the key to differential privacy's success, as it permits the construction of complex – and useful – differentially private analyses from simple differentially private primitives. That is, it allows us to "program" in a privacy-preserving fashion. Optimizing for composition, and being willing to live with high probability bounds on privacy loss rather than insisting on worst-case bounds, led to the introduction of *concentrated differential privacy* (CDP), a guarantee incomparable to the others, permitting better accuracy than both without compromising on cumulative privacy loss under composition [6, 14]. Intuitively, CDP requires that the privacy loss have small expectation, and tight concentration around its expectation. There is no hard threshold for maximal privacy loss. Instead, the probability of large losses vanishes at a rate that roughly parallels the concentration of the Gaussian distribution. This framework exactly captures the type of privacy loss that occurs under composition of (pure and approximate) DP mechanisms. CDP improves on approximate DP by "cutting corners" in a way that has no privacy cost under high levels of composition. For certain approximate DP algorithms, such as the Gaussian mechanism, when a δ-probability failure occurs, there is no privacy catastrophe. Rather, the most likely situation is that the privacy loss is still bounded by 2ε. More generally, for any $k \geq 1$, the risk that the privacy loss exceeds $k\varepsilon$ is $\delta^{\Omega(k^2)}$. In such situations, the $\sqrt{\log(1/\delta)}$ multiplicative loss in accuracy one obtains in the analysis of this algorithm is a high price to pay to avoid a small increase in privacy loss. Under high levels of composition of approximate DP mechanisms, the privacy losses of each individual mechanism are forced to be "artificially"

STOC'18, June 25–29, 2018, Los Angeles, CA, USA

ACM ISBN 978-1-4503-5559-9/18/06…$15.00
https://doi.org/10.1145/3188745.3188946

low in order to control the cumulative loss. This means that small lapses in these artificially low losses are not so consequential, and literally do not add up to much in the risk of exceeding the desired cumulative loss.

Finally, in addition to providing better accuracy than pure and approximate DP, CDP has a simple optimal composition theorem, computable in linear time. This stands in sharp contrast to the situation with approximate DP, for which computing optimal composition bounds is computationally intractable (#P-hard) [20].

With these important advantages, concentrated DP also has limitations:

(1) "Packing"-based lower bounds [6, 16] imply inherent limitations on the accuracy that can be obtained under CDP for certain tasks. These lower bounds do not apply to approximate DP (because of the δ-probability "escape hatch"), and there are tasks where the accuracy of CDP data analyses is worse than what can be obtained under approximate DP.
(2) Unlike both pure and approximate DP, CDP does not support "privacy amplification" via subsampling [17, 22]. This type of privacy amplification is a powerful tool, as demonstrated by the recent work of Abadi *et al.* in differentially private deep learning [1]. While their privacy analysis utilizes a CDP-like view of privacy loss, their use of subsampling for privacy amplification means that their algorithm cannot be analyzed within the framework of CDP.

With the above progress and challenges in mind, we propose the notion of *truncated concentrated differential privacy* (tCDP) to give us the best of all words.

A new framework. Loosely speaking, CDP restricts the privacy loss to be at least as concentrated as a Gaussian. tCDP relaxes this requirement. Informally, it only requires Gaussian-like concentration up to a set boundary, specified by a number of standard deviations. Beyond this boundary, the privacy loss can be less concentrated than a Gaussian (but we still require *subexponential* concentration). The formal definition uses two parameters: ρ, which (roughly) controls the expectation and standard deviation of the privacy loss, and ω, which (roughly) controls the number of standard deviations for which we require Gaussian-like concentration. The privacy guarantee is stronger the smaller ρ is and the larger ω is.

The formal definition uses Rényi divergences to restrict the privacy loss random variable:

Definition 1. *Let $\rho > 0$ and $\omega > 1$. A randomized algorithm $M : \mathcal{X}^n \to \mathcal{Y}$ satisfies* ω-truncated ρ-concentrated differential privacy (or (ρ, ω)-tCDP) *if, for all neighboring $x, x' \in \mathcal{X}^n$,*[1]

$$\forall \alpha \in (1, \omega)\ \mathrm{D}_\alpha\left(M(x) \middle\| M(x')\right) \leq \rho\alpha,$$

where $\mathrm{D}_\alpha(\cdot\|\cdot)$ denotes the Rényi divergence [21, 24] of order α (in nats, rather than bits).[2]

The definition of tCDP is a direct relaxation of zero-concentrated differential privacy (zCDP) [6] and is closely related to Rényi differential privacy [19]. In particular, setting $\omega = \infty$ exactly recovers the definition of ρ-zCDP. Clearly, decreasing ω relaxes the definition of (ρ, ω)-tCDP. See Section 1.3 for further discussion of the relationship between tCDP and prior definitions.

CDP: a recap. The addition of Gaussian noise drawn from $\mathcal{N}\left(0, \left(\frac{\Delta f}{\varepsilon}\right)^2\right)$ is a universal mechanism for Concentrated Differential Privacy, yielding a privacy loss random variable Z that is subgaussian with standard $\lambda = \varepsilon$, and with expected privacy loss $\mu = \varepsilon^2/2$. Formally, we have the following bound on the moment generating function:

$$\forall \lambda \in \mathbb{R}\ \mathbb{E}[e^{\lambda(Z-\mu)}] \leq e^{\varepsilon^2\lambda^2/2}. \tag{1}$$

This implies that $\Pr[Z - \mu \geq k\varepsilon] \leq e^{-k^2/2}$, where the connection between bounds on high moments and tail bounds goes via an application of Markov's inequality:

$$\Pr[(Z-\mu) \geq k\varepsilon] = \Pr[e^{(\lambda/\varepsilon)(Z-\mu)} \geq e^{\lambda k}] \leq \frac{\mathbb{E}[e^{(\lambda/\varepsilon)(Z-\mu)}]}{e^{\lambda k}}. \tag{2}$$

This is minimized when $\lambda = k/\varepsilon$, yielding a probability bounded by $e^{-k^2/2}$.

Subsampling under CDP and tCDP.. Given a dataset of size n, let us choose a random subset of size sn, for $s \in (0, 1)$, as a preprocessing step before applying a differentially private algorithm. In analogy to the situation for pure and approximate differential privacy, we would expect subsampling to decrease the privacy loss roughly by a factor of s. More precisely we would hope that the new standard should be $O(s\varepsilon)$, with the mean reduced by a factor of roughly s^2. Let us see what actually happens.

Let x be a dataset of n zeroes, and let y be a neighboring dataset of $n-1$ zeroes and a single one. Consider the sensitivity 1 query q, "How many 1's are in the dataset?" Let Z be a random variable with distribution $\mathcal{N}(0, \left(\frac{1}{\varepsilon}\right)^2)$. Let A be the event that the response received is at least $1 + z$, where $\Pr[Z \geq z] \leq e^{-k^2/2}$ for a fixed $k > 0$. Letting $\sigma^2 = \left(\frac{\Delta q}{\varepsilon}\right)^2 = 1/\varepsilon^2$, when the dataset is y the output event A corresponds to the sampling event $Z \geq k\sigma$, and when the dataset is x the output event A corresponds to the sampling event $Z \geq (k + \varepsilon)\sigma$. A rough calculation immediately yields:

$$\frac{\Pr[A|y]}{\Pr[A|x]} \approx \frac{e^{-k^2/2}}{e^{-(k+\varepsilon)^2/2}} = e^{\varepsilon k + \varepsilon^2/2}. \tag{3}$$

This corresponds to a privacy loss of εk above the mean ($\varepsilon^2/2$). Let q_x denote the probability of output event A after subsampling and noise addition, when the dataset is x and the probability is taken both over the subsampling and the draw from the Gaussian, and let q_y be defined analogously. Then $q_x = p_x$ because the subsampled dataset (of course) still contains only zeroes and so again the probability of the output event is the same as the probability that the noise exceeds $(k + \varepsilon)\sigma$. However, q_y is more complicated: with probability s the single 1 is in the subsample, in which case the output event corresponds to the noise event $Z \geq k\sigma$, and with probability $(1 - s)$ the subsample does not capture the 1, in which case the output event corresponds to the noise event $Z \geq (k + \varepsilon)\sigma$.

[1] x and x' are neighboring (denoted $x \sim x'$) if they differ on a single entry – i.e. only one person's data is changed.

[2] The Rényi divergence of order α between two distributions P and Q over a sample space Ω (with $P \ll Q$) is defined to be $\mathrm{D}_\alpha(P\|Q) = \frac{1}{\alpha-1}\log\int_\Omega P(y)^\alpha Q(y)^{1-\alpha}\mathrm{d}y$. Here log denotes the natural logarithm.

Letting $\eta = e^{-k^2/2}$ we get:

$$\frac{q_y}{q_x} \approx \frac{s\eta + (1-s)\eta e^{-\varepsilon k} e^{-\varepsilon^2/2}}{\eta e^{-\varepsilon k} e^{-\varepsilon^2/2}} \approx s e^{\varepsilon k} + (1-s).$$

When $\varepsilon k < 1$ this is (roughly) $e^{s\varepsilon k}$, corresponding to a privacy loss of $s\varepsilon k$ above the (new) mean, as desired. But when $k \geq 1/\varepsilon$ the constraint is violated; thus Concentrated Differential Privacy does not fully benefit from subsampling and the subgaussian standard does not decrease from ε to $s\varepsilon$. With this in mind tCDP does the next best thing: it ensures that we get the desired reduction for all k less than a fixed bound of approximately $\omega\varepsilon$. In the case of the (subsampled) Gaussian, this bound is roughly $1/\varepsilon$. For this reason we parameterize tCDP with the expected privacy loss, usually denoted ρ, and an upper bound, usually denoted ω, that controls the moments for which the subgaussian property must apply. Specifically, tCDP guarantees that we can indeed bound the probability on the left hand side of Equation 2 by $e^{-k^2/2}$ for all $k \leq \varepsilon(\omega - 1)$. (The "$-1$" comes from a technicality that arises in the translation from bounding the Rényi divergence to bounding the moment generating function.)

The most closely related work in the literature is the moments accountant technique of [1], which analyzes the integer positive moments of the Gaussian mechanism under subsampling. They obtain similar bounds to those obtained here, although they collapse higher moments into an error term, δ, and present their results in terms of (ε, δ)-differential privacy. Our bounds apply to all moments and all mechanisms ensuring Concentrated Differential Privacy. Moreover, by adhering to the framework of CDP we are able to circumvent the #P-hardness of computing optimal composition bounds.

Group privacy. Concentrated Differential Privacy automatically yields group privacy, replacing the standard ε with the new standard $g\varepsilon$ for a group of size g, and so bounding the privacy loss for a group of size g corresponds to bounding the gth moment of the original privacy loss random variable. (ρ, ω)-tCDP enjoys the same property, but only for groups of size at most $\varepsilon\omega$, since that is the largest moment for which the bound in Equation 2 is required to hold. Naturally, the bounds on the moments of the *group privacy loss* random variable only hold for $\lambda \leq \varepsilon(\omega - 1)/g$ (otherwise we could "unroll" this bundling into groups and bound higher order moments beyond ω).

Summary of Results. We introduce and perform a detailed study of tCDP. Beyond our introduction of this new framework for privacy-preserving data analysis, our contributions include:

- **Robust guarantees.** We show optimal and efficiently computable composition bounds for tCDP. Examining group privacy, we show strong bounds for groups of size at most $O(\sqrt{\rho} \cdot \omega)$. For larger groups, we show weaker but nontrivial bounds. These bounds for large groups also enable us to prove "packing-based" error lower bounds that show our results are tight.
- **Privacy amplification via subsampling.** We show that the privacy guarantees of tCDP can be amplified by subsampling. In a nutshell, sampling an s-fraction of the individuals in the dataset before running a (ρ, ω)-tCDP mechanism gives $(O(\rho s^2), \omega')$-tCDP, where ω' is close to the original ω. Privacy amplification is a powerful technique in data-rich regimes (see e.g. [1]). We note that for CDP, subsampling does not improve the privacy parameters. The proof that tCDP is amplified by subsampling proceeds via a delicate analysis of the privacy loss random variable, and is one of our primary technical contributions.
- **A tCDP toolkit.** We construct tCDP mechanisms for basic data analysis tasks. These mechanisms, together with tight composition bounds and privacy amplification via subsampling, provide a rich toolkit for tCDP data analysis. First, we introduce a canonical noise-adding mechanism for answering low-sensitivity queries. To guarantee tCDP, we use noise drawn from a novel "sinh-normal" distribution. The tails of this noise distribution are exponentially tighter than a Gaussian (so the probability of large errors is exponentially smaller). Other basic tCDP tools we introduce include adding Gaussian noise with data-dependent variance, and a mechanism for answering point queries with improved accuracy.
- **The power of tCDP.** Building on the new tools described above, we turn to more advanced data analysis tasks. As our main additional contribution, we consider the following "Gap-Max" task capturing optimization over a discrete set. Given a collection of k low-sensitivity queries and a real parameter GAP > 0, we want to identify the query whose answer is maximal, under the promise that its answer is at least GAP larger than the second-largest answer. We want as weak a promise as possible, i.e. to minimize GAP (as a function of k). We show a tCDP algorithm when GAP is as small as (roughly) $O(\log \log k)$. This is an exponential improvement over what is possible using CDP (or pure DP), and can be a significant improvement over the bounds obtainable for approximate DP when δ is cryptographically small.
  The Gap-Max optimization task has many applications in the design of privacy-preserving algorithms. We show it implies tCDP algorithms for releasing histograms over k bins and for answering threshold queries, with improved accuracy bounds in both cases (see below).

Organization. The remainder of the Introduction is organized as follows. Properties of tCDP flowing from the definition are spelled out in **Section 1.1**. In **Section 1.2** we describe our results on tCDP mechanisms: building blocks and more advanced algorithms. **Section 1.3** provides a detailed discussion of the qualitative and quantitative relationships between tCDP and earlier variants of differential privacy. For the remainder of the paper, **Section 2** contains additional definitions and an analysis of group privacy for groups of size greater than ω. **Section 3** contains the details on amplification via subsampling. **Section 4** discusses the sinh-normal noise distribution and shows that the addition of noise from an appropriately scaled version of this distribution is a canonical mechanism for achieving tCDP. **Section 5** provides several applications of tCDP using more sophisticated algorithms than simple noise addition. This section contains our above-mentioned results on the Gap-Max problem. Finally, **Section 6** shows the optimality of our algorithmic results by proving matching lower bounds.

1.1 Properties of tCDP

Several important guarantees follow from the definition of tCDP.

Composition and postprocessing. tCDP provides tight and tractable composition bounds, and is also robust to postprocessing:

Lemma 2 (Composition & Postprocessing). *Let $M : \mathcal{X}^n \to \mathcal{Y}$ satisfy (ρ, ω)-tCDP. Let $M' : \mathcal{X}^n \times \mathcal{Y} \to \mathcal{Z}$ be such that $M'(\cdot, y) : \mathcal{X}^n \to \mathcal{Z}$ satisfies (ρ', ω')-tCDP for all $y \in \mathcal{Y}$. Define $M'' : \mathcal{X}^n \to \mathcal{Z}$ by $M''(x) = M'(x, M(x))$. Then M'' satisfies $(\rho + \rho', \min\{\omega, \omega'\})$-tCDP.*

Group privacy. tCDP protects the privacy of groups of size at most $\varepsilon\omega$. In particular, (ρ, ω)-tCDP for individuals implies $(\rho \cdot g^2, \omega/g)$-tCDP for groups of size g, and this is tight.

For larger groups, tCDP still provides gracefully degrading privacy protection, but the rate of degradation is more rapid (exponential, rather than quadratic; Section 2.2, Proposition 9). In comparison, for (ε, δ)-differential privacy, privacy is guaranteed for groups of size at most $O(\log(1/\delta)/\varepsilon)$, with no protection at all for larger groups. As it turns out, the weaker protections afforded to large groups can be a blessing in disguise, as the packing lower bounds that bedevil CDP are consequently substantially weaker for tCDP (see Section 6).

Privacy amplification. We show that the privacy guarantees of tCDP are amplified when sampling a s-fraction of the dataset entries:

Proposition 3 (Privacy Amplification by Subsampling). *Let $M : \mathcal{X}^n \to \mathcal{Y}$ satisfy (ρ, ω)-tCDP. Define $M' : \mathcal{X}^N \to \mathcal{Y}$ to be $M'(x) = M(x_S)$ where $S \subset [N]$ with $|S| = n$ is uniformly random and, $x_S \in \mathcal{X}^n$ is the restriction of $x \in \mathcal{X}^N$ to the entries specified by S. Let $s = n/N$ be the subsampling fraction. Then M' satisfies $(O(\rho s^2), \Omega(\min\{\omega, \log(1/s)/\rho\}))$-tCDP.*

Intuitively, for an initial ω that wasn't too large (say $\omega \le 1/\rho$), subsampling maintains $\omega' = \Omega(\omega)$. Up to constant factors, this result is optimal. In contrast, subsampling *cannot* improve the guarantees of a CDP mechanism: s-fraction subsampling of a ρ-CDP algorithm is still at best ρ-CDP. Meanwhile, approximate (ε, δ)-differential privacy with s-fraction subsampling yields $(\hat{\varepsilon}, s\delta)$-differential privacy with $\hat{\varepsilon} = \log(1 + s(e^\varepsilon - 1)) \approx s\varepsilon$ (e.g., [5]).

1.2 tCDP Mechanisms

What useful algorithmic techniques are compatible with tCDP? Of course, all techniques that are compatible with CDP and pure differential privacy immediately carry over to tCDP. We show that additional tools, which capture much of the power of approximate differential privacy, are amenable to analysis within the framework of tCDP.

Additive noise. Perhaps the most basic challenge to consider is answering a single low-sensitivity query $q : \mathcal{X}^n \to \mathbb{R}$ subject to tCDP. Let Δ be the global sensitivity of q, i.e. $\max_{x \sim x'} |q(x) - q(x')| \le \Delta$. For DP and its previously-studied relaxations, the canonical mechanisms for this basic task use additive noise, drawn from the appropriate distribution (e.g. Laplace for pure DP or Gaussian for CDP), scaled to the query's sensitivity and to the privacy parameter. Similarly, we propose the "sinh-normal" as a canonical noise distribution for tCDP:

Proposition 4 (Sinh-Normal Mechanism). *Let $q : \mathcal{X}^n \to \mathbb{R}$ have sensitivity Δ. Define the* sinh-normal mechanism*, with parameters $\rho > 0$ and $\omega > 1/\sqrt{\rho}$, to be the additive noise mechanism that on input $x \in \mathcal{X}^n$ releases $q(x) + Z$, where*

$$Z \leftarrow \omega \cdot \Delta \cdot \operatorname{arsinh}\left(\frac{1}{\omega \cdot \Delta} \cdot \mathcal{N}(0, \Delta^2/2\rho)\right). \tag{4}$$

This mechanism satisfies $(O(\rho), O(\omega))$-tCDP.

Here, the inverse hyperbolic sine function — $\operatorname{arsinh}(y) = \log(y + \sqrt{y^2 + 1})$ — is an even sigmoidal function, with $\operatorname{arsinh}(y) \approx y$ for $|y| \approx 0$ and $\operatorname{arsinh}(y) \approx \operatorname{sign}(y) \cdot \log|y|$ for $|y| \gg 0$. Hence, the sinh-normal distribution closely resembles a Gaussian with comparable variance up to noise values of magnitude roughly $(\omega \cdot \Delta)$. Beyond this boundary, the tails decay exponentially faster than a Gaussian, roughly as $e^{-e^{t\sqrt{\rho}/\Delta}}$, where t is the noise magnitude and ω is taken to be as small as possible.

The sinh-normal that guarantees (ρ, ω)-tCDP and the Gaussian that guarantees ρ-CDP are similar for noise values of magnitude smaller than $(\omega \cdot \Delta)$ For larger noise values the tails of the sinh-normal are exponentially tighter, which results in better accuracy than the Gaussian. While this difference may not be crucial when answering a single query (since these are low-probability events), it becomes important when one needs to bound the maximum of many independent samples from the sinh-normal distribution.

Stable selection. Optimizing over a discrete set is a basic task for privacy-preserving algorithms. We consider the following formulation of this problem: we are given a collection of sensitivity-1 functions $q_1, \ldots, q_k : \mathcal{X}^n \to \mathbb{R}$, a sensitive dataset $x \in \mathcal{X}^n$, and a real parameter GAP > 0. Let $i^* = \operatorname{argmax}_{i \in [k]} q_i(x)$. The *stable selection problem* is to identify i^* with high probability under the promise that

$$q_{i^*}(x) \ge \max_{i \in [k] \setminus \{i^*\}} q_i(x) + \text{GAP}\,.$$

That is, we are promised that the largest $q_i(x)$ is at least GAP larger than the second largest. Weaker notions of privacy permit smaller values of GAP, which in turn is a less stringent condition on how stable the maximizer needs to be to perform selection. Focusing on the dependence of GAP on the number of choices k, we know that $(\varepsilon, 0)$-differential privacy requires $\text{GAP} = \Theta(\log(k)/\varepsilon)$, whereas ρ-zCDP requires $\text{GAP} = \Theta(\sqrt{\log(k)/\rho})$ and (ε, δ)-differential privacy allows for $\text{GAP} = \Theta(\log(1/\delta)/\varepsilon)$ — independent of k.

We show a (tight) bound of $\text{GAP} = \Theta(\omega \cdot \log\log k)$ for stable selection under tCDP:

Proposition 5. *Let $\rho \in (0, 1)$ and $\omega \ge 1/\sqrt{\rho}$. Then there is a (ρ, ω)-tCDP algorithm for stable selection as long as*

$$\text{GAP} \ge O\left(\omega \cdot \log\left(1 + \frac{\sqrt{\log k}}{\omega\sqrt{\rho}}\right)\right).^{3}$$

[3] Note that if $\omega \gtrsim \sqrt{\log(k)/\rho}$, then this expression becomes $\sqrt{\log(k)/\rho}$, matching the bound for ρ-zCDP.

This result demonstrates an exponential savings over CDP and pure differential privacy. This bound is incomparable to the $\Theta(\log(1/\delta)/\varepsilon)$ achieved by (ε, δ)-differential privacy. While the $\log\log k$ term introduces a (mild) dependence on the number of functions, it may often be dominated by $\log(1/\delta)$ for reasonable choices of k and δ. Stable selection is an extremely versatile primitive that already captures the stability arguments used in many prior works in approximate differential privacy. As an example application, we can use stable selection to design a tCDP algorithm for releasing a histogram over k bins with ℓ_∞ error $O(\log\log k)$. Similarly, we can answer threshold queries with error $O(\log\log k)$. A packing argument using weak group privacy (Corollary 27) shows that Proposition 5 is optimal.

We prove Proposition 5 as a consequence of a generic construction that allows us to convert any additive noise mechanism into a solution to the stable selection problem, while preserving whatever notion of privacy the original noise mechanism satisfies. The required gap corresponds to the tail behavior of the additive noise. In particular, we recover all of the aforementioned results by applying this transformation with Laplace noise, Gaussian noise, and truncated Laplace noise, respectively.

1.3 Relationship to Prior Definitions

The differences between the various flavors of differential privacy — pure differential privacy, approximate differential privacy, concentrated differential privacy, and tCDP — come down to how they treat very low probability events. That is, for neighboring inputs x and x' and a set of outcomes S, how are $\mathbb{P}[M(x) \in S]$ and $\mathbb{P}[M(x') \in S]$ related when both are small?

Pure $(\varepsilon, 0)$-differential privacy applies the same strict multiplicative guarantee to all events, regardless of how small their probability. Meanwhile, approximate (ε, δ)-differential privacy simply "ignores" events with probability less than δ. CDP, tCDP, and Rényi differential privacy (RDP) fall somewhere in between; no events are completely ignored, but the requirements placed on low probability events significantly relax those of pure differential privacy.

A helpful way to characterize these definitions is in terms of tail bounds on the privacy loss. Recall that the privacy loss of a randomized algorithm M for neighboring inputs x and x' is defined to be the random variable $Z = f(M(x))$ where $f(y) = \log(\mathbb{P}[M(x) = y]/\mathbb{P}[M(x') = y])$.

- Pure $(\varepsilon, 0)$-differential privacy requires that the privacy loss is bounded: $Z \le \varepsilon$.

- ρ-zCDP requires that Z is subgaussian — that is, the tail behaviour of Z should be like that of $\mathcal{N}(\rho, 2\rho)$, with $\mathbb{P}[Z > t + \rho] \le e^{-t^2/(4\rho)}$ for all $t \ge 0$.

- (ρ, ω)-tCDP also requires Z to be subgaussian near the origin, but only subexponential in its tails. That is, as with ρ-zCDP, we have $\mathbb{P}[Z > t + \rho] \le e^{-t^2/(4\rho)}$ for $0 \le t \le 2\rho(\omega - 1)$. But for $t > 2\rho(\omega - 1)$, we get a weaker subexponential tail bound of the form $\mathbb{P}[Z > t + \rho] \le e^{(\omega-1)^2\rho} \cdot e^{-(\omega-1)t}$.

- (ω, τ)-RDP [19] requires $\mathrm{D}_\alpha(M(x)\|M(x')) \le \tau$ for all $\alpha \le \omega$. Recall that (ρ, ω)–tCDP ensures that $\mathrm{D}_\alpha(M(x)\|M(x')) \le \rho\alpha$ for all $\alpha \le \omega$. It is helpful to compare the definitions when $\tau = \rho\omega$. In this case both RDP and tCDP ensure $\mathrm{D}_\alpha(M(x)\|M(x')) \le \rho\omega$ for all $\alpha \le \omega$. However, tCDP *in addition* requires $\mathrm{D}_\alpha(M(x)\|M(x')) \le \rho\alpha$ for all α in the interval $(1, \omega)$.

  In terms of tail bounds, RDP requires the privacy loss Z to be subexponential. That is, $\mathbb{P}[Z > t + \tau] \le e^{-(\omega-1)t}$. The definition of tCDP requires a similar subexponential tail bound, but in addition requires subgaussian behavior for values of t that are not too large.

- Up to constant factors, (ε, δ)-differential privacy is equivalent to requiring $\mathbb{P}[Z > \varepsilon] \le \delta$.

- The notion δ-approximate ρ-zCDP [6] was proposed as a unification of CDP and approximate DP. This definition requires there to be an event E occuring with probability at least $1 - \delta$ such that the privacy loss conditioned on E is subgaussian. While this definition permits simple composition and stable selection algorithms, its behavior under subsampling is less clear.

Since (ρ, ∞)-tCDP is equivalent to ρ-zCDP, $(\varepsilon, 0)$-differential privacy implies $(\frac{1}{2}\varepsilon^2, \infty)$-tCDP [6, 14]. Conversely, tCDP implies a family of (ε, δ)-differential privacy guarantees:

Lemma 6. *Suppose M satisfies (ρ, ω)-tCDP. Then, for all $\delta > 0$ and all $1 < \alpha \le \omega$, M satisfies (ε, δ)-differential privacy with*

$$\varepsilon = \begin{cases} \rho + 2\sqrt{\rho\log(1/\delta)} & \text{if } \log(1/\delta) \le (\omega-1)^2\rho \\ \rho\omega + \frac{\log(1/\delta)}{\omega-1} & \text{if } \log(1/\delta) \ge (\omega-1)^2\rho \end{cases}.$$

Parameter settings. With the above comparisons in mind, one can think of the ρ parameter as analogous to the value $\frac{1}{2}\varepsilon^2$ for a pure $(\varepsilon, 0)$-DP (or approximate (ε, δ)-DP for sufficiently small δ) algorithm, and similar to the ρ parameter in CDP. As discussed above, the truncation point ω controls the maximal moment for which the bound in Equation 2 is required to hold, which in turn specifies the maximal size of groups that receive meaningful group privacy protection.

Adopting the convention that "meaningful" privacy corresponds to $\varepsilon = 1$, pure ε-DP gives meaningful privacy protections for groups of size $1/\varepsilon$. For CDP and tCDP this corresponds to groups of size $\sqrt{1/2\rho} = (1/\varepsilon)$. Recalling that tCDP protects groups of size roughly $\sqrt{\rho} \cdot \omega$, an apt choice for ω would thus be anything greater or equal to roughly $1/\sqrt{2}\rho$.

This choice of ω provides a comparable level of protection to that of CDP for events that don't have tiny probabilities. We note that the probabilities of larger privacy losses still vanish exponentially quickly (as compared with the subgaussian guarantee of CDP).

We note that privacy amplification via subsampling reduces ρ to a smaller value ρ', but results in a new $\omega' \approx \omega$ that is roughly unchanged (see Proposition 3). After subsampling it can be the case that $\omega' \ll 1/\rho'$. This is reasonable in the context of composition, where many mechanisms with very small privacy losses are composed, and the primary concern is the *global* privacy loss under this composition. Recall that while the ρ parameter degrades under composition, ω does not (see Lemma 2). Thus, the value of ω for

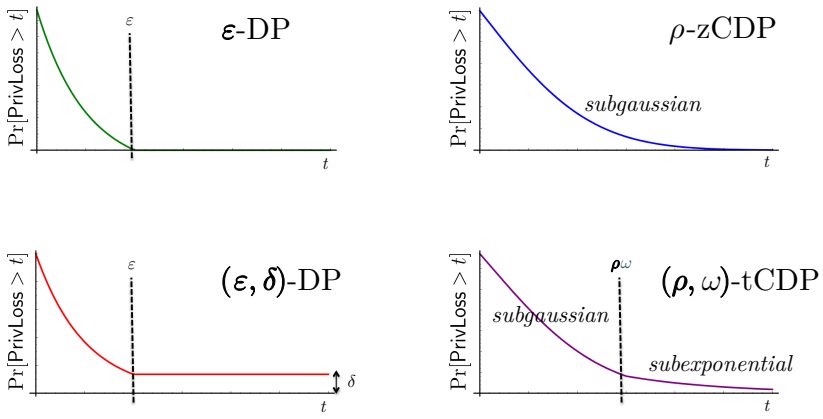

Figure 1: Caricatures of privacy loss tail bounds permissible under each formulation of differential privacy.

each individual mechanism can be set with an eye towards the total privacy loss under composition. Subsampling will not improve ω, but neither will composition degrade it.

2 DEFINITION AND BASIC PROPERTIES

We will work with the following definition.

Definition 7. *A mechanism $M : \mathcal{X}^n \to \mathcal{Y}$ satisfies ω-truncated ρ-concentrated differential privacy (abbreviated (ρ, ω)-tCDP) if, for all $x, x' \in \mathcal{X}^n$ differing in a single entry,*

$$\forall \alpha \in (1, \omega)\ \mathrm{D}_\alpha\left(M(x) \middle\| M(x')\right) \le \rho\alpha.$$

Here $\mathrm{D}_\alpha(\cdot\|\cdot)$ denotes the Rényi divergence of order α. For distributions P and Q on Ω it is defined by

$$\mathrm{D}_\alpha(P\|Q) = \frac{\log\left(\int_\Omega P(y)^\alpha Q(y)^{1-\alpha}\,\mathrm{d}y\right)}{\alpha - 1}$$

where $P(\cdot)$ and $Q(\cdot)$ represent the probability density/mass functions of P and Q respectively. (More generally, $P(y)/Q(y)$ is the Radon-Nikodym derivate of P with respect to Q evaluated at y.) Note that we use log to denote the natural logarithm.

Composition and postprocessing are clean and simple for this definition (Lemma 2). Below, we examine other properties of this definition.

2.1 Conversion to DP

Lemma 6 shows that tCDP implies a (ε, δ)-DP guarantee for every $\delta > 0$. This provides one way of interpreting the guarantee of tCDP.

Lemma 8. *Let P and Q be probability distributions on Ω. Suppose $\mathrm{D}_\alpha(P\|Q) \le \rho\alpha$. Then, for every event E and all $\delta > 0$, we have $P(E) \le e^\varepsilon Q(E) + \delta$ for $\varepsilon = \rho\alpha + \log(1/\delta)/(\alpha - 1)$.*

Note that the optimal choice of α (assuming the choice is unconstrained) is $\alpha - 1 = \sqrt{\log(1/\delta)/\rho}$, which yields $\varepsilon = \rho + 2\sqrt{\rho\log(1/\delta)}$. To attain Lemma 6, we set $\alpha = \min\{\omega, 1 + \sqrt{\log(1/\delta)/\rho}\}$.

2.2 Group Privacy

The proof of group privacy for zCDP [6] immediately extends to show that (ρ, ω)-tCDP yields $(\rho k^2, \omega/k)$-tCDP for groups of size k. However, that proof does not yield any results for $k > \omega$. Here, we develop a different bound which applies to larger groups.

Proposition 9 (Group Privacy for Large Groups). *Let $M : \mathcal{X}^n \to \mathcal{Y}$ satisfy (ρ, ω)-tCDP. Let $x, x' \in \mathcal{X}^n$ differ in k entries. Then*

$$\mathrm{D}_\alpha\left(M(x)\middle\|M(x')\right) \le \left(\left(1 + \frac{1}{\omega - 1}\right)^k - 1\right)(\omega - 1)\omega\rho \le e^{\frac{k}{\omega-1}} \cdot \omega^2 \cdot \rho$$

for

$$\alpha = 1 + \frac{1}{\left(1 + \frac{1}{\omega-1}\right)^k - 1} \ge 1 + e^{\frac{-k}{\omega-1}} > 1.$$

The group privacy properties of tCDP for groups of size $k > \omega$ are substantially weaker than those of CDP: First, the bound grows exponentially in the group size. Second, we only obtain bounds on $\mathrm{D}_{1+o(1)}(M(x)\|M(x'))$, which approaches the KL divergence.

To help provide further intuition about this group privacy guarantee, we describe the tail bound (i.e. (ε, δ)-differential privacy) guarantee that it entails. Roughly speaking, we obtain an (ε, δ)-differential privacy guarantee for every $\delta > 0$ where ε depends *exponentially* on the group size k, and logarithmically on $1/\delta$.

Proposition 10 (Interpreting Group Privacy). *Let $M : \mathcal{X}^n \to \mathcal{Y}$ satisfy (ρ, ω)-tCDP. Let $x, x' \in \mathcal{X}^n$ differ in k entries. Then for every $\delta > 0$ and every measurable $S \subseteq \mathcal{Y}$ we have $\Pr[M(x) \in S] \le e^\varepsilon \Pr[M(x') \in S] + \delta$ where*

$$\varepsilon = e^{\frac{k}{\omega-1}}\left(\omega^2\rho + \log(1/\delta)\right).$$

2.3 Basic Algorithms

Most of the basic tools of differential privacy are compatible with CDP and, hence, with tCDP. In particular, the Laplace [12] and Exponential [18] mechanisms satisfy $(\varepsilon, 0)$-DP, which means they satisfy $\frac{1}{2}\varepsilon^2$-zCDP [6, Prop. 1.4] and hence $(\frac{1}{2}\varepsilon^2, \infty)$-tCDP. Adding Gaussian noise sampled from $\mathcal{N}(0, \sigma^2)$ to a sensitivity-Δ query

guarantees $\left(\frac{\Delta^2}{2\sigma^2}\right)$-zCDP [6, Prop. 1.6] and $\left(\frac{\Delta^2}{2\sigma^2}, \infty\right)$-tCDP. Later we will show that tCDP supports additional algorithmic techniques.

3 PRIVACY AMPLIFICATION BY SUBSAMPLING

A more complex property of tCDP is privacy amplification by subsampling. This is a property that CDP does *not* provide. Informally, this property states that, if a private algorithm is run on a random subset of a larger dataset (and the identity of that subset remains hidden), then this new algorithm provides better privacy protection (reflected through improved privacy parameters) to the entire dataset as a whole than the original algorithm did.

Theorem 11 (Privacy Amplification by Subsampling). *Let $\rho, s \in (0, 0.1]$ and $n, N \in \mathbb{N}$ with $s = n/N$ and $\log(1/s) \geq 3\rho(2 + \log_2(1/\rho))$. Let $M : \mathcal{X}^n \to \mathcal{Y}$ satisfy (ρ, ω')-tCDP for $\omega' \geq \frac{1}{2\rho} \cdot \log(1/s) \geq 3$. Define the mechanism $M_s : \mathcal{X}^N \to \mathcal{Y}$ by $M_s(x) = M(x_S)$, where $x_S \in \mathcal{X}^n$ is the restriction of $x \in \mathcal{X}^N$ to the entries specified by a uniformly random subset $S \subseteq [N]$ with $|S| = n$.*

The algorithm $M_s : \mathcal{X}^N \to \mathcal{Y}$ satisfies $(13s^2\rho, \omega)$-tCDP for

$$\omega = \frac{\log(1/s)}{4\rho}.$$

Before delving into the proof we compare this result to the analogous property of DP: Let $M : \mathcal{X}^n \to \mathcal{Y}$ and let $M_s : \mathcal{X}^N \to \mathcal{Y}$ be defined as in Theorem 11 with $s = n/N$. If M satisfies (ε, δ)-DP, then M_s satisfies $(\log(1+s\cdot(e^\varepsilon-1)), s\cdot\delta)$-DP. Note that $\log(1+s\cdot(e^\varepsilon-1)) \approx s\cdot\varepsilon$. (In somewhat more detail, $s\cdot\varepsilon \leq \log(1+s\cdot(e^\varepsilon-1)) \leq s\cdot\varepsilon + \frac{\varepsilon^2}{8}$ by Hoeffding's lemma.)

To conduct the proof, we first set up some notation. Let $M : \mathcal{X}^N \to \mathcal{Y}$ satisfy ρ-zCDP. Fix $x, x' \in \mathcal{X}^N$ differing in a single index $i \in [n]$. Let $S \subset [N]$ with $|S| = n$ be uniformly random and let $x_S, x'_S \in \mathcal{X}^n$ denote the restrictions of x and x', respectively, to the indices in S. We define the following probability densities:

$$\begin{aligned} P &= \text{the density of } M(x_S) \text{ conditioned on } i \in S, \\ Q &= \text{the density of } M(x'_S) \text{ conditioned on } i \in S, \\ R &= \text{the density of } M(x_S) \text{ conditioned on } i \notin S, \\ &= \text{the density of } M(x'_S) \text{ conditioned on } i \notin S. \end{aligned}$$

Since index i appears in the sample S with probability s, we may write the density of $M_s(x) = M(x_S)$ as $sP + (1-s)R$. Similarly, the density of $M_s(x') = M(x'_S)$ is $sQ + (1-s)R$.

Since $M : \mathcal{X}^n \to \mathcal{Y}$ is (ρ, ω')-zCDP (for some ω'), from the quasi-convexity of Rényi divergence, it holds that $\mathrm{D}_{\alpha'}(P_1 \| P_2) \leq \rho\alpha'$ for every pair $P_1, P_2 \in \{P, Q, R\}$ and every $\alpha \in (1, \omega')$. Therefore, in order to prove Theorem 11, it suffices to prove the following result about Rényi divergences. The remainder of this section is devoted to proving Theorem 12.

Theorem 12. *Let $\rho \in (0, 0.1]$, $s \in (0, 1]$, and $\omega' > 2$ satisfy*

$$2 + \log_2\left(\frac{1}{\rho}\right) + \log_2\left(\frac{1}{(\omega'-2)s}\right) \leq \omega' \leq \frac{1}{\rho} \cdot \log\left(\frac{1}{(\omega'-2)s}\right). \quad (5)$$

Let P, Q, R be probability density functions over $\mathcal{Y}$ such that

$$\mathrm{D}_\alpha(P_1 \| P_2) \leq \rho\alpha$$

for every pair $P_1, P_2 \in \{P, Q, R\}$ and all $\alpha \in (1, \omega')$. Then

$$\mathrm{D}_\alpha(sP + (1-s)R \,\|\, sQ + (1-s)R) \leq 13s^2\rho\alpha$$

for every $\alpha \in (1, \omega)$, where $\omega = \omega'/2$.

Under the mild constraint that

$$\log(1/s) \geq 3\rho(2 + \log_2(1/\rho)),$$

Condition 5 holds as long as

$$\omega' = \frac{1}{2\rho} \cdot \log(1/s) \geq 3.$$

These values are used in the statement of Theorem 11.

We begin by stating some auxiliary lemmata, which follow from elementary calculus.

Lemma 13. *Let $\alpha > 1$ and $-1 \leq x \leq 1/(\alpha-1)$. Then $(1+x)^\alpha \leq 1 + \alpha x + \alpha(\alpha-1)x^2$.*

Lemma 14. *For $x \geq -1$, $\alpha \in [1, \infty)$, and $\beta \geq 0$,*

$$(1+x)^\alpha \leq 1 + \alpha x + \alpha(\alpha-1)x^2 + (\alpha x_0)^\alpha((\alpha-1)x_0)^\beta,$$

where $x_0 = \max\{x, 0\}$.

Proof Sketch of Theorem 12. For notational convenience, let $U = sQ + (1-s)R$. By quasi-convexity and continuity of Rényi divergence, note that $\mathrm{D}_\alpha(P\|U) \leq \rho\alpha$ and $\mathrm{D}_\alpha(Q\|U) \leq \rho\alpha$ for all $\alpha \in [1, \omega']$. Fix $\alpha \in (1, \omega)$. We begin by writing

$$\begin{aligned} e^{(\alpha-1)\mathrm{D}_\alpha(sP+(1-s)R \,\|\, sQ+(1-s)R)} &= e^{(\alpha-1)\mathrm{D}_\alpha(sP+(1-s)R \,\|\, U)} \\ &= \mathop{\mathbb{E}}_{y\leftarrow U}\left[\left(\frac{sP(y) + (1-s)R(y)}{U(y)}\right)^\alpha\right] \\ &= \mathop{\mathbb{E}}_{y\leftarrow U}\left[\left(1 + s\cdot\left(\frac{P(y)-Q(y)}{U(y)}\right)\right)^\alpha\right] \\ &= \mathbb{E}\left[(1+sZ)^\alpha\right] \end{aligned}$$

where $Z = (P(y) - Q(y))/U(y)$ for $y \leftarrow U$.

Observe that $\mathbb{E}[Z] = 0$. We now estimate $\mathbb{E}\left[Z^2\right]$. To do this, we make use of the polarization identity

$$(P-Q)^2 = 2(P^2+Q^2) - 4\left(\frac{P}{2} + \frac{Q}{2}\right)^2.$$

This lets us write

$$\begin{aligned} \mathbb{E}\left[Z^2\right] &= 2e^{\mathrm{D}_2(P\|U)} + 2e^{\mathrm{D}_2(Q\|U)} - 4\exp\left(\mathrm{D}_2\left(\frac{1}{2}P + \frac{1}{2}Q\middle\|U\right)\right) \\ &\leq 2e^{2\rho} + 2e^{2\rho} - 4 = 4(e^{2\rho} - 1). \end{aligned}$$

Here, the inequality holds because $\mathrm{D}_2(P\|U), \mathrm{D}_2(Q\|U) \leq 2\rho$ and Rényi divergence is nonnegative.

Define $Z_0 = \max\{Z, 0\}$ and $\hat{Z} = P(y)/U(y)$ for $y \leftarrow U$. Note that $Z_0 \leq \hat{Z}$. Since $1 + sZ$ is obtained as a ratio of probabilities, $sZ \geq -1$. Therefore, we may apply Lemma 14 to estimate

$$\begin{aligned} \mathbb{E}\left[(1+sZ)^\alpha\right] &\leq \mathbb{E}\left[1 + \alpha sZ + \alpha(\alpha-1)s^2Z^2 + (\alpha sZ_0)^\alpha((\alpha-1)sZ_0)^\beta\right] \\ &= 1 + \alpha s\cdot\mathbb{E}[Z] + \alpha(\alpha-1)s^2\cdot\mathbb{E}\left[Z^2\right] + (\alpha s)^\alpha((\alpha-1)s)^\beta\cdot\mathbb{E}\left[Z_0^{\alpha+\beta}\right] \\ &\leq 1 + 4s^2\alpha(\alpha-1)\cdot(e^{2\rho}-1) + (\alpha s)^\alpha((\alpha-1)s)^\beta\mathbb{E}\left[\hat{Z}^{\alpha+\beta}\right] \\ &\leq 1 + 4s^2\alpha(\alpha-1)\cdot(e^{2\rho}-1) + (\alpha s)^\alpha((\alpha-1)s)^\beta\cdot e^{(\alpha+\beta)(\alpha+\beta-1)\rho}. \end{aligned}$$

The penultimate inequality holds because of our estimates of $\mathbb{E}[Z]$ and $\mathbb{E}\left[Z^2\right]$, and the fact that $0 \le Z_0 \le \hat{Z}$. The final inequality follows because $\mathrm{D}_{\alpha+\beta}(P\|U) \le \rho(\alpha+\beta)$ as long as $\alpha+\beta \le \omega'$.

We now estimate the term on the right, taking $\beta = \omega' - \alpha \ge 0$. Here, we have

$$\begin{aligned}(\alpha s)^{\alpha}\cdot((\alpha-1)s)^{\omega'-\alpha}\cdot\exp(\omega'(\omega'-1)\rho)\\ \le e\alpha(\alpha-1)s^2\cdot((\alpha-1)s)^{\omega'-2}\cdot\exp(\omega'(\omega'-1)\rho)\end{aligned}$$

Condition 5 reveals that for $1 < \alpha \le \omega'/2$, we have

$$((\alpha-1)s)^{\omega'-2}\cdot\exp(\omega'(\omega'-1)\rho) \le \rho.$$

We may now complete the calculation with

$$\begin{aligned}\mathbb{E}\left[(1+sZ)^{\alpha}\right] &\le 1 + 4\alpha(\alpha-1)s^2(e^{2\rho}-1) + e\alpha(\alpha-1)s^2\cdot\rho\\ &\le 1 + 13\alpha(\alpha-1)s^2\rho \le e^{\alpha(\alpha-1)\cdot 13s^2\rho},\end{aligned}$$

as required. □

3.1 Optimality of our Subsampling Bound

We next show that our subsampling bound (Theorem 11) is optimal up to constants. To do this we use the following result which uses the binomial theorem to give an exact bound for privacy amplification by subsampling (but only for the special case where α is an integer, and the distributions Q and R are identical).

Lemma 15 (Exact Subsampling). *Let P and Q be probability distributions and $s \in [0,1]$. Suppose $\alpha \in (1,\infty)$ is an integer. Then*

$$\mathrm{D}_{\alpha}(sP+(1-s)Q\|Q) = \frac{1}{\alpha-1}\log\left((1-s)^{\alpha} + \alpha(1-s)^{\alpha-1}s + \sum_{k=2}^{\alpha}\binom{\alpha}{k}(1-s)^{\alpha-k}s^k e^{(k-1)\mathrm{D}_k(P\|Q)}\right).$$

Fix $\rho, s \in (0, 0.1]$. To show the optimality of Theorem 12, consider $P = \mathcal{N}(1, 1/2\rho)$ and $Q = R = \mathcal{N}(0, 1/2\rho)$. This corresponds to the Gaussian Mechanism, which satisfies (ρ,∞)-tCDP. In fact, $\mathrm{D}_{\alpha}(P\|Q) = \rho\alpha$ for all $\alpha \in (1,\infty)$.

Theorem 12 implies $\mathrm{D}_{\alpha}(sP+(1-s)Q\|Q) \le 13s^2\rho\alpha$ for all $\alpha \in (1,\omega)$, where $\omega = \min\{\log(1/s)/3\rho + 1/2, 1 + 1/s^{1/3}\}$.

By Lemma 15 with $\alpha = 2$,

$$\begin{aligned}\mathrm{D}_{\alpha}(sP+(1-s)Q\|Q) &= \log\left((1-s)^2 + 2(1-s)s + s^2 e^{\mathrm{D}_2(P\|Q)}\right)\\ &= \log\left(1 + s^2(e^{2\rho}-1)\right)\\ &\ge s^2\rho\alpha.\end{aligned}$$

This shows that, while we may be able to reduce the constant 13 in our bound to 1, we cannot reduce it further.

Next we show the optimality of ω. Theorem 12 asserts that $\mathrm{D}_{\alpha}(sP+(1-s)Q\|Q) \le 10s^2\rho\alpha$ for all $\alpha \in (1,\omega)$. We will consider the largest value of ω for which this holds. Let $\alpha = \lceil\omega - 1\rceil \ge 2$. By Lemma 15,

$$\begin{aligned}\mathrm{D}_{\alpha}(sP+(1-s)Q\|Q) &\ge \frac{1}{\alpha-1}\log\left(s^{\alpha}e^{(\alpha-1)\mathrm{D}_{\alpha}(P\|Q)}\right)\\ &= \rho\alpha - \frac{\alpha}{\alpha-1}\log(1/s).\end{aligned}$$

Now we have

$$10s^2\rho\alpha \ge \mathrm{D}_{\alpha}(sP+(1-s)Q\|Q) \ge \rho\alpha - \frac{\alpha}{\alpha-1}\log(1/s),$$

which entails

$$\omega - 1 \le \alpha \le \frac{1}{1-10s^2}\cdot\frac{\log(1/s)}{\rho}.$$

For sufficiently small s, this shows that $\omega = O(\log(1/s)/\rho)$ is necessary. This matches our bound up to constant factors.

4 SINH-NORMAL NOISE

One of the most basic techniques in differential privacy is the use of noise addition to answer a single low-sensitivity query. We recall the definition of (global) sensitivity.

Definition 16. *Let $q : \mathcal{X}^n \to \mathbb{R}$. The global sensitivity of q is the minimum $\Delta > 0$ such that $|q(x) - q(x')| \le \Delta$ for all $x, x' \in \mathcal{X}^n$ differing on a single entry.*

A low-sensitivity query q may be answered with differential privacy by adding noise with standard deviation proportional to Δ. Each variant of differential privacy (pure, approximate, concentrated, truncated concentrated) provides constraints on the distributions from which this noise may be drawn. For example, pure differential privacy requires a noise distribution whose tails decay at most inverse exponentially. The canonical distribution with this property is the Laplace distribution.

We can similarly identify "canonical" distributions for approximate and concentrated differential privacy. For (ε,δ)-differential privacy, we believe this to be the Laplace distribution with standard deviation Δ/ε, but with its support truncated to the interval $[\pm O(\Delta\log(1/\delta)/\varepsilon)]$. For CDP, the canonical noise distribution is the Gaussian.

For tCDP, we propose adding noise sampled from the following distribution, for parameters $\sigma, A > 0$.

$$X \leftarrow A\cdot\operatorname{arsinh}\left(\frac{\sigma}{A}\cdot\mathcal{N}(0,1)\right).$$

Here $\operatorname{arsinh}(x) = \log\left(x + \sqrt{x^2+1}\right)$ is the inverse of the hyperbolic sine function $\sinh(y) = \frac{1}{2}(e^y - e^{-y})$. Intuitively, X is simply the Gaussian $\mathcal{N}(0,\sigma^2)$ with exponentially faster tail decay. To see this, note that around the origin $\operatorname{arsinh}(x) \approx x$, so the distribution looks like a Gaussian here. However, when x is large, $\operatorname{arsinh}(x) \approx \log x$. (Since arsinh is a symmetric function, the lower tails are identical.) Thus the tails decay doubly exponentially, rather than just in a subgaussian manner. The value of A determines where the transition from linear to logarithmic occurs; as $A \to \infty$, we have $X \to \mathcal{N}(0,\sigma^2)$. The quantity A also controls the truncation point in the resulting tCDP guarantee.

Theorem 17. *Let $q : \mathcal{X}^n \to \mathbb{R}$ have sensitivity-Δ. Let ρ, A satisfy $1 < 1/\sqrt{\rho} \le A/\Delta$. Define a randomized algorithm $M : \mathcal{X}^n \to \mathbb{R}$ by*

$$M(x) \leftarrow q(x) + A\operatorname{arsinh}\left(\frac{1}{A}\mathcal{N}(0,\Delta^2/2\rho)\right).$$

Then M satisfies $(16\rho, A/8\Delta)$-tCDP.

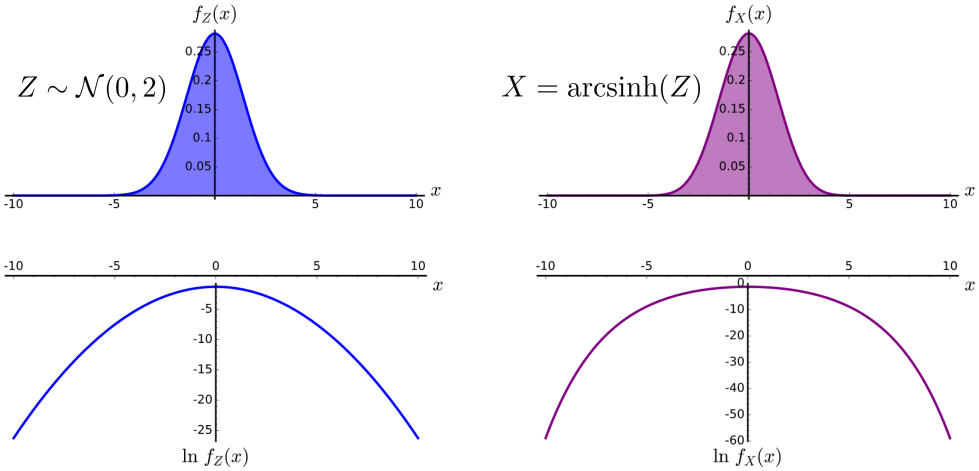

Figure 2: Left: The probability density function f_Z of $Z \leftarrow \mathcal{N}(0,2)$ and its logarithm. Right: The probability density function f_X of the sinh-normal distribution $X = \operatorname{arsinh}(Z)$ and its logarithm.

We remark that the bound of Theorem 17 is optimal up to constant factors. (Namely, strengthening Theorem 17 would correspondingly strengthen our result for point queries in Corollary 18, which would violate the lower bound given in Corollary 27.)

The proof of Theorem 17 requires some calculus, and appears in the full version of this work.

4.1 Example Application: Histograms

To demonstrate the advantage of the sinh-normal noise over Gaussian noise, we consider the example of privately releasing histograms. That is, for each $u \in \mathcal{X}$ there is a point query $q_u : \mathcal{X}^n \to \mathbb{R}$ given by $q_u(x) = |\{i \in [n] : x_i = u\}|$. That is, $q_u(x)$ counts the number of times u appears in x.

Under $(\varepsilon, 0)$-differential privacy, we can, with high probability, answer all queries with error $\Theta\left(\frac{\log|\mathcal{X}|}{\varepsilon}\right)$. Under (ε,δ)-differential privacy this can potentially be improved to $\Theta\left(\frac{\log(1/\delta)}{\varepsilon}\right)$. For ρ-zCDP the error scales as $\Theta\left(\sqrt{\frac{\log|\mathcal{X}|}{\rho}}\right)$. In all cases, these error bounds are attained by adding independent noise from the Laplace, truncated Laplace, and Gaussian distributions respectively. We now show that adding noise from the sinh-normal distribution attains error $O(\omega \log\log|\mathcal{X}|)$:

Corollary 18. *For every $\rho \in (0,1)$ and $\omega \geq 1/8\sqrt{\rho}$, there exists a (ρ,ω)-tCDP randomized algorithm $M : \mathcal{X}^n \to \mathbb{R}^{\mathcal{X}}$ such that for every $\beta > 0$ and $x \in \mathcal{X}^n$,*

$$\mathop{\mathbb{P}}_{M}\left[\max_{u\in\mathcal{X}} |M(x)_u - q_u(x)| \geq 8\omega \cdot \operatorname{arsinh}\left(\sqrt{\frac{\log(|\mathcal{X}|/\beta)}{2\omega^2\rho}}\right)\right] \leq \beta.$$

If $\frac{\log(|\mathcal{X}|/\beta)}{2\omega^2\rho} \approx 0$, then $8\omega\cdot\operatorname{arsinh}\left(\sqrt{\frac{\log(|\mathcal{X}|/\beta)}{2\omega^2\rho}}\right) \approx 4\sqrt{\frac{2\log(|\mathcal{X}|/\beta)}{\rho}}$, as for $x \approx 0$ we have $\operatorname{arsinh}(x) \approx x$. This matches the bound attainable under ρ-zCDP up to constant factors.

If $\frac{\log(|\mathcal{X}|/\beta)}{2\omega^2\rho} \gg 0$, then

$$8\omega \cdot \operatorname{arsinh}\left(\sqrt{\frac{\log(|\mathcal{X}|/\beta)}{2\omega^2\rho}}\right) \approx 4\omega \cdot \left(\log\log(|\mathcal{X}|/\beta) + \log(2/\omega^2\rho)\right),$$

as for $x \gg 0$ we have $\operatorname{arsinh}(x) \approx \log(2x)$. This $\log\log|\mathcal{X}|$ dependence is exponentially better than what is attainable under CDP. Later (Corollary 27), we show that this bound is tight up to constant factors.

Proof. We can answer these queries by adding independent sinh-normal noise to each query. Clearly, each q_u has sensitivity $\Delta = 1$. Furthermore, if $x, x' \in \mathcal{X}^n$ differ in a single coordinate, then for all but two of the q_u queries we have $q_u(x) = q_u(x')$. Hence, the ℓ_1 sensitivity of the entire vector of queries is 2, so as we will see, it suffices to add independent sinh-normal noise at scale $\Delta = 2$ to every individual query.

More precisely, we define an algorithm M as follows. Independently for each $u \in \mathcal{X}$, we sample

$$M(x)_u \leftarrow q_u(x) + 8\omega \cdot \operatorname{arsinh}\left(\frac{1}{8\omega} \cdot \mathcal{N}(0, 16/\rho)\right).$$

By Theorem 17, every individual answer $M(x)_u$ satisfies $(\rho/2,\omega)$-tCDP. Fix $x, x' \in \mathcal{X}$ differing in a single entry. For all but two points $u \in \mathcal{X}$, we have $q_u(x) = q_u(x')$, so the privacy loss corresponding to every such coordinate is zero. Applying composition over the two remaining coordinates yields the desired privacy guarantee.

The accuracy guarantee follows from a union bound and the fact that

$$\forall \sigma, x > 0 \quad \mathop{\mathbb{P}}_{X\leftarrow\mathcal{N}(0,\sigma^2)}[|X| \geq x] \leq e^{-x^2/2\sigma^2}.$$

□

5 SOPHISTICATED ALGORITHMS

We have covered the key algorithmic primitives of tCDP – all those compatible with CDP plus sinh-normal noise addition and Gaussian

smooth sensitivity. Now we demonstrate how these primitives can be used to build more sophisticated algorithms.

We define the Gap-Max problem, which is a useful primitive for differentially private data analysis:

Problem 19 (Gap-Max). *Our problem is specified by a loss function $\ell : \mathcal{X}^n \times [k] \to \mathbb{R}$ and a gap $\alpha \geq 0$. We assume ℓ has sensitivity 1 in its first argument – that is, for all neighbouring $x, x' \in \mathcal{X}^n$, we have*

$$\forall j \in [k] \; |\ell(x, j) - \ell(x', j)| \leq 1.$$

We seek a differentially private algorithm $M : \mathcal{X}^n \to [k]$ with the following accuracy guarantee.

Fix an input $x \in \mathcal{X}^n$. Let

$$j_* = \operatorname{argmax}_{j \in [k]} \ell(x, j).$$

Suppose

$$\forall j \in [k] \; j \neq j_* \implies \ell(x, j) < \ell(x, j_*) - \alpha n.$$

Under this "gap assumption," we want $M(x) = j_*$ with high probability.

Gap-Max is a useful primitive that has been studied extensively in the literature (both explicitly and implicitly). As we show below, it is also an informative test-case for studying different variants of differential privacy. The first algorithm for solving this problem is the exponential mechanism [18], defined by

$$\mathbb{P}[M(x) = j] \propto \exp\left(\frac{\varepsilon}{2} \cdot \ell(x, j)\right).$$

This algorithm is $(\varepsilon, 0)$-DP (and $(\frac{1}{2}\varepsilon^2, \infty)$-tCDP) and solves the gap-max problem as long as $\alpha n \geq O(\log(k)/\varepsilon)$. (This bound is optimal, up to constants, for pure DP. More generally, the exponential mechanism is optimal for the "approx-max" problem where there is no gap assumption.)

However, we can do better than the exponential mechanism if we relax to approximate DP. The propose-test-release framework [11] can be used to solve the gap-max problem subject to (ε, δ)-DP as long as $\alpha n \geq O(\log(1/\delta)/\varepsilon)$. (Note that this bound is independent of k.) What can be done with tCDP?

We give two generic reductions from privacy-preserving mechanisms that answer low-sensitivity queries by adding (appropriately sampled) noise to the gap-max problem. The two reductions obtain similar parameters (up to constant factors). Both reductions give rise to resulting mechanisms that attain whatever differential privacy guarantee the original noise-adding mechanism attained. The required gap depends on the tail behaviour of the noise distribution. A somewhat-informal statement follows:

Theorem 20 (Additive Noise to Gap-Max Reduction). *Let $\mathcal{D}$ be a distribution on $\mathbb{R}$ such that the following hold.*

- *Adding noise sampled from $\mathcal{D}$ to a sensitivity-1 query satisfies (a certain variant of) differential privacy.*
- $\mathbb{P}_{X \leftarrow \mathcal{D}}[|X| \geq \alpha n/2] \ll 1/k$.

Then there exists a differentially private algorithm solving the gap-max problem with gap α and k options. The attained differential privacy guarantee is the same as that afforded by the noise-addition mechanism (up to postprocessing and two-fold composition).

Applying our reduction to the Laplace distribution attains a $(\varepsilon, 0)$-DP algorithm for gap-max as long as $\alpha n \geq O(\log(k)/\varepsilon)$. Applying our reduction to the Laplace distribution with support restricted to a symmetric interval around the origin of length $O(\log(1/\delta)/\varepsilon)$ attains a (ε, δ)-DP algorithm for gap-max as long as $\alpha n \geq O(\log(\min\{k, 1/\delta\}/\varepsilon)$. Applying our reduction to the Gaussian distribution attains a ρ-zCDP algorithm for gap-max as long as $\alpha n \geq O(\sqrt{\log(k)/\rho})$. Finally, applying our reduction to the sinh-normal distribution attains a (ρ, ω)-tCDP algorithm for gap-max as long as $\alpha n \geq O\left((\omega + 1/\sqrt{\rho}) \cdot \log\log k\right)$. See Section 1.2 for a discussion comparing the parameters obtained under the various variants of differential privacy.

5.1 First Reduction

Our first reduction replaces the loss function $\ell : \mathcal{X}^n \times [k] \to \mathbb{R}$ with a surrogate function f that is "sparse" over $[k]$ (much like a histogram). Specifically, we construct a function $f : \mathcal{X}^n \times [k] \to \mathbb{R}$ such that for every $x \in \mathcal{X}^n$, the functions f and ℓ have the same argmax (over $[k]$) and the same gap. Moreover, $f(x, j)$ has small sensitivity (in x, for fixed j), and f is sparse: for each database $x \in \mathcal{X}^n$, the value of $f(x, \cdot)$ is 0 on all but a single coordinate $j^* \in [k]$. The combination of these properties allows us find the argmax of f (while satisfying the appropriate variant of differential privacy) by simply adding noise to each of the k values and returning the noisy argmax. The privacy argument is similar to the approach used for releasing histograms (or, equivalently, answering point queries). The argmax of f equals the argmax of ℓ, and thus whenever this procedure succeeds (w.r.t f) it correctly identifies ℓ's argmax.

Algorithm 21. ***Generalized Histogram Algorithm***

- `Parameters:` $n, k \in \mathbb{N}$ `and a continuous distribution` $\mathcal{D}$ `on` $\mathbb{R}$.
- `Public input:` $\ell : \mathcal{X}^n \times [k] \to \mathbb{R}$ `such that` $\ell(\cdot, j)$ `has sensitivity 1 for all` $j \in [k]$.
- `Private input:` $x \in \mathcal{X}^n$
- `Define a permutation` $\sigma_x : [k] \to [k]$ `such that` $\ell(x, \sigma_x(j)) \geq \ell(x, \sigma_x(j+1))$ `for all` $j \in [k-1]$. `I.e. the permutation` σ_x `orders the elements of` $[k]$ `in descending order according to their loss.`
- `(Goal: Output` $\sigma_x(1) = \operatorname{argmax}_{j \in [k]} \ell(x, j) \in [k]$, `subject to differential privacy.)`
- `Define` $f(x, j) = \max\{\ell(x, j) - \ell(x, \sigma_x(2)), 0\}$.
- `Independently for each` $j \in [k]$ `sample` $Z_j \leftarrow \mathcal{D}$.
- `Output` $\hat{j} = \operatorname{argmax}_{j \in [k]} f(x, j) + Z_j$.

The key observation is that, for any x, $f(x, j) = 0$ for all $j \neq \sigma_x(1)$. In other words, f is very sparse. Furthermore, for fixed $j \in [k]$, f has sensitivity 2. Thus, for neighbouring inputs x, x', if we look at the sensitivity vector $v = (f(x, 1) - f(x', 1), f(x, 2) - f(x', 2), \cdots, f(x, k) - f(x', k))$, we have that all but (at most) two entries in v equal 0 ($\|v\|_0 \leq 2$) and $\|v\|_\infty \leq 2$. This means that the privacy cost of our mechanism is the same as answering up to 2 queries of sensitivity at most 2 with the noise distribution $\mathcal{D}$.

On the utility front, the gap assumption implies that $f(x, \sigma_x(1)) \geq \alpha n$ and $f(x, j) = 0$ for all $j \neq \sigma_x(1)$. We just need to union bound over all the noise to ensure that the noisy max is the true max

with high probability. That is, if $|Z_j| \le \alpha n/2$ for all $j \in [k]$, then $\hat{j} = \sigma_x(1)$.

This algorithm readily generalizes to the gap version of the "top-m problem" (more commonly called top-k) which has been studied extensively [2, 4, 7, 23]: We simply set $f(x,j) = \max\{\ell(x,j) - \ell(x,\sigma_x(m+1)), 0\}$ instead. Now the sparsity is m (rather than 1). So we pay a $\sqrt{m}$ factor from composition and are able to identify up to m indices as long as their values are sufficiently larger than the $(m+1)^{\text{th}}$ value.

5.2 Second Reduction

Our second reduction is a bit sharper: we reduce gap-max to a *single* instance of the noise-addition mechanism (rather than two instances as in Section 5.1). The reduction is a bit more complex (and assumes that the noise added is symmetric). This reduction operates by adding noise to the actual gap in the given database. If the noisy gap is larger than a threshold, then the (true) argmax is returned. Otherwise, the reduction returns one of the other (incorrect) elements chosen uniformly and at random. We note that this closely mimics the propose-test-release framework [11].

The threshold is chosen carefully so that both privacy and accuracy are satisfied. For accuracy, we want the threshold to be large enough. For privacy, since the gap is a low-sensitivity quantity, the noisy gap itself is differentially private (and by post-processing, privacy is maintained when we compare the noisy gap to a fixed threshold). The twist comes in that when the noisy gap is large enough, we release the true argmax (which depends directly on the data, and is thus not a differentially private quantity). To guarantee privacy, we carefully choose the threshold so for any two neighboring databases x and x', either their argmax is the same (in which case privacy follows immediately), or otherwise both their gaps are quite small, and in this case the output distribution on both databases is uniformly random. The details follow:

Algorithm 22. ***Obvious Answer Mechanism***

- Parameters: $n, k \in \mathbb{N}$ and a continuous distribution $\mathcal{D}$ on $\mathbb{R}$.
- Public input: $\ell : \mathcal{X}^n \times [k] \to \mathbb{R}$ such that $\ell(\cdot, j)$ is sensitivity-1 for all $j \in [k]$.
- Private input: $x \in \mathcal{X}^n$
- Define a permutation $\sigma_x : [k] \to [k]$ such that $\ell(x, \sigma_x(j)) \ge \ell(x, \sigma_x(j+1))$ for all $j \in [k-1]$.
- (Goal: Output $\sigma_x(1) = \text{argmax}_{j \in [k]}\ \ell(x,j) \in [k]$, subject to differential privacy.)
- Let $\gamma(x) = \ell(x, \sigma_x(1)) - \ell(x, \sigma_x(2)) \in \mathbb{R}$.
- Let $\hat{\gamma}(x) = \max\left\{\frac{1}{2} \cdot \gamma(x) - 1, 0\right\}$.
- Compute a threshold $t_k(\mathcal{D}) \in \mathbb{R}$ such that $\mathbb{P}_{Z \leftarrow \mathcal{D}}[Z > t_k(\mathcal{D})] = 1/k$. (Exact equality is important.)
- Sample $Z \leftarrow \mathcal{D}$.
- If $\hat{\gamma}(x) + Z > t_k(\mathcal{D})$, output $\sigma_x(1)$; else output $\sigma_x(j)$ for a uniformly random $j \in \{2, 3, \cdots, k\}$.

Clearly, if $\mathcal{D}$ is a symmetric distribution and $\hat{\gamma}(x) \ge 2t_k(\mathcal{D})$ or, equivalently, $\gamma(x) \ge 4t_k(\mathcal{D}) + 2$, then $\mathbb{P}[M(x) = \sigma_x(1)] \ge 1 - 1/k$. This is our accuracy guarantee – if the gap is sufficiently large, the correct answer is returned with high probability. The threshold $t_k(\mathcal{D})$ is determined by the tails of the distribution $\mathcal{D}$. In particular, by the guarantees of Theorem 20, we know that $t_k(\mathcal{D}) \ll \alpha n/2$, and thus the algorithm succeeds w.h.p. when the gap is at least αn.

Lemma 23. *Suppose $\mathcal{D}$ is a differentially private noise-adding distribution for sensitivity-1 queries. Then the obvious answer mechanism is differentially private with exactly the same privacy guarantee. (This only assumes that the particular variant of differential privacy is closed under postprocessing and that an algorithm that ignores its input is differentially private.)*

Proof. Fix neighbouring inputs $x, x' \in \mathcal{X}^n$.

Claim: Either $\sigma_x(1) = \sigma_{x'}(1)$ or $\gamma(x) + \gamma(x') \le 2$.

Before proving the claim, we shall complete the rest of the proof using a case analysis based on the claim.

Case 1: $\sigma_x(1) = \sigma_{x'}(1)$.

In this case we can treat the value of $\sigma_x(1) = \sigma_{x'}(1)$ as a constant when comparing the output distributions. Now the only private computation is deciding whether or not to output this constant. Since $\hat{\gamma}(\cdot)$ is a sensitivity-1 function and we assume that $\mathcal{D}$ is a differentially private noise-adding distribution, the value of $\hat{\gamma}(\cdot) + Z$ is a differentially private output. By postprocessing, the decision $\hat{\gamma}(\cdot) + Z > t(\mathcal{D})$ is differentially private and so is the output of the mechanism.

Case 2: $\gamma(x) + \gamma(x') \le 2$.

In this case $\gamma(x) \le 2$ and $\hat{\gamma}(x) = \max\{\frac{1}{2} \cdot \gamma(x) - 1, 0\} = 0$. Thus $\mathbb{P}[\hat{\gamma}(x) + Z > t_k(\mathcal{D})] = 1/k$. Hence $\mathbb{P}[M(x) = \sigma_x(1)] = 1/k$. For $j > 1$, we also have

$$\begin{aligned}\mathbb{P}[M(x) = \sigma_x(j)] &= \mathbb{P}[\hat{\gamma}(x) + Z \le t_k(\mathcal{D})] \cdot \frac{1}{k-1} \\ &= (1 - 1/k)/(k-1) = 1/k.\end{aligned}$$

Thus $M(x)$ outputs a uniformly random element from $[k]$. Likewise, $M(x')$ outputs a uniformly random answer. Since both distributions of $M(x)$ and $M(x')$ are identical, we trivially satisfy the differential privacy constraint.

Proof of Claim: Let $j = \sigma_x(1)$ and $j' = \sigma_{x'}(1)$. Suppose $j \ne j'$. We must show that $\gamma(x) + \gamma(x') \le 2$. By the definition of $j = \sigma_x(1)$ and $\gamma(x)$, we have $\ell(x,j) - \gamma(x) \ge \ell(x,j')$. Likewise, $\ell(x',j') - \gamma(x') \ge \ell(x',j)$. Since ℓ is low-sensitivity in its first argument, $\ell(x,j') \ge \ell(x',j') - 1$ and $\ell(x',j) \ge \ell(x,j) - 1$. We now add up these four inequalities to obtain

$$\begin{pmatrix} \ell(x,j) - \gamma(x) \\ +\ell(x',j') - \gamma(x') \\ +\ell(x,j') \\ +\ell(x',j) \end{pmatrix} \ge \begin{pmatrix} \ell(x,j') \\ +\ell(x',j) \\ +\ell(x',j') - 1 \\ +\ell(x,j) - 1 \end{pmatrix},$$

which simplifies to $\gamma(x) + \gamma(x') \le 2$. □

5.3 Threshold Queries

We next use our gap-max algorithm as a subroutine for answering threshold queries over a totally ordered domain. While thresholds are extremely simple functions, the problem of releasing answers to threshold queries has received significant attention and has led to the development of a number of new algorithmic techniques (e.g., [3, 5, 13]). For a domain size k, the threshold queries $q_1, \cdots, q_k : [k]^n \to \mathbb{R}$ are defined by $q_j(x) = |\{i \in [n] : x_i \le j\}|$.

As with point queries, the problem of releasing k threshold queries has sample complexity $n = \Theta(\log(k)/\varepsilon)$ for pure $(\varepsilon, 0)$-DP and $n = \Theta(\sqrt{\log(k)/\rho})$ for ρ-zCDP. However, for approximate (ε, δ)-DP, the sample complexity's dependence on the number of thresholds k is much more intriguing. While it lies between $\log^* k$ and $2^{\log^* k}$ [3, 5], the true dependence on k remains an open problem.

In this section, we show that the sample complexity of releasing thresholds with (ρ, ω)-tCDP is $n = \tilde{O}((\omega + 1/\sqrt{\rho}) \cdot \log\log k)$, where the $\tilde{O}$ notation hides factors sublogarithmic in $\log\log k$. To do so, we leverage a reduction from [5] which proved a tight relationship between the sample complexity of answering thresholds and that of a simpler problem called the *interior point problem*. Given a dataset $x \in [k]^n$, a solution to the interior point problem is a number j such that $\min_{i\in[n]} x_i \le j \le \max_{i\in[n]} x_i$. That is, j lies (non-strictly) between the least element and the greatest element of the dataset x. The reduction of [5] showed that if the interior point problem has sample complexity n, then threshold queries can be answered to accuracy α using sample complexity $n \cdot \tilde{O}(1/\alpha)$.

We actually describe our solution to the interior point problem under tCDP as the solution to a more general problem. The generalized formulation lends itself more naturally to our recursive construction.

Problem 24 (Generalized Interior Point). *Let $f : \mathcal{X}^n \times [k] \to [0, 1]$ be a function such that*

- *The function f has sensitivity Δ in its first argument. That is, for all $x, x' \in \mathcal{X}^n$ which differ in a single entry, we have $|f(x, j) - f(x', j)| \le \Delta$.*
- *The function f is nondecreasing in its second argument. That is, $f(x, j) \le f(x, j + 1)$ for all $1 \le j \le k - 1$. For notational convenience, define $f(x, 0) = 0$ and $f(x, k + 1) = 1$ for all x.*

We seek a differentially private algorithm $M : \mathcal{X}^n \to [k + 1]$ such that the following accuracy guarantee holds with high probability: If $j = M(x)$, then $0 < f(x, j)$ and $f(x, j - 1) < 1$.

To see that Problem 24 generalizes the interior point problem, observe that we recover the formulation of the interior point problem by simply taking $f(x, j) = q_j(x) = |\{i \in [n] : x_i \le j\}|$, which has sensitivity $\Delta = 1$.

Theorem 25. *The generalized interior point problem with sensitivity Δ can be solved with (ρ, ω)-tCDP with sample complexity $n = 2^{O(\log^* k)} \cdot (\omega + 1/\sqrt{\rho}) \cdot \Delta \cdot \log\log k$.*

In this extended abstract, we provide a high-level description of our recursive algorithm for the generalized interior point problem. Each level of recursion reduces an instance of the problem with dataset size n, domain size k, and sensitivity Δ to a problem with dataset size n, domain size $O(\log k)$, and sensitivity $O(\Delta)$. After $O(\log^* k)$ levels of recursion, we get a problem with constant domain size and sensitivity $2^{O(\log^* k)}\Delta$ which can be solved directly, e.g., with the exponential mechanism.

We now describe one level of recursion of the algorithm. For simplicity, assume that $k + 1$ is a power of two. Given an input dataset $x \in \mathcal{X}^n$, the algorithm first constructs a complete binary tree with $k + 1$ leaves and depth $\log_2(k + 1) + 1$. Each node of the tree corresponds to a dyadic subinterval of $[k + 1]$. That is, every leaf node corresponds to a single domain item, and every internal node corresponds to the interval between its leftmost descendant and its rightmost descendant. The *value* of a node corresponding to an interval $[j_1, j_2]$ is the amount by which the function $f(x, \cdot)$ increases over that interval, i.e., $f(x, j_2) - f(x, j_1)$.

This tree gives rise to a smaller instance of the generalized interior point problem by inducing a function $\hat{f} : \mathcal{X}^n \times [\log_2(k + 1) + 1]$, where $\hat{f}(x, \hat{j})$ is the maximum value of any node at level $\hat{j}$ of the tree (numbering levels in increasing order from leaves to root). Given a solution to the generalized interior point problem for $\hat{f}$, we now argue how to combine it with the gap-max algorithm to solve the original problem for f.

Supposing our recursive call succeeds, we obtain an answer $\hat{j}$ such that there exists a node at level $\hat{j}$ with value greater than 0, but every node at level $\hat{j} - 1$ has value less than 1. A small technical modification lets us make the stronger guarantee that some node at level $\hat{j}$ actually has value greater than $1/5$, while every node at level $\hat{j} - 1$ has value less than $2/5$. Using the gap-max algorithm, we can identify a node at level $\hat{j}$, corresponding to some interval $[j_1, j_2]$, such that $f(x, j_2) - f(x, j_1) > 0$.

On the other hand, since every node at level $\hat{j} - 1$ has value less than $2/5$, we have $f(x, j_2) - f(x, j_1) < 4/5$. Hence, either $1/10 < f(x, j_1) < 1$ or $0 < f(x, j_2) < 9/10$. A simple differentially private test then allows us to identify at least one of the endpoints j_1, j_2 which serves as a generalized interior point.

6 LOWER BOUNDS

CDP, like pure DP and unlike approximate DP, is subject to information theoretic "packing" lower bounds [6, 8, 16]. (Although for CDP these are quadratically weaker than for pure DP.) We now prove packing lower bounds for tCDP. These lower bounds are *exponentially* weaker than the corresponding ones for CDP and pure DP. Having weaker lower bounds is a good thing – provided we don't open the door for privacy violations.

The basis for our lower bounds is the following mutual information bound.

Theorem 26. *Let $M : \mathcal{X}^n \to \mathcal{Y}$ be a randomized algorithm satisfying (ρ, ω)-tCDP. Let X be a random variable on $\mathcal{X}^n$. Then*

$$I(X; M(X)) \le e^{\frac{n}{\omega-1}} \omega^2 \rho,$$

where $I(\cdot;\cdot)$ denotes the mutual information (in nats).

Note that if $\omega = n + 1$, then, up to a constant factor, this recovers the bound $I(X; M(X)) \le n^2\rho$ which holds for ρ-zCDP [6, Thm. 1.10] and is the basis for CDP packing lower bounds.

Proof. By the group privacy property of tCDP (Proposition 9), we have that for any $x, x' \in \mathcal{X}^n$ (i.e. we allow these to differ on n entries, corresponding to group size $k = n$),

$$\mathrm{D}_\alpha\left(M(x) \middle\| M(x')\right) \le \left(\left(1 + \frac{1}{\omega - 1}\right)^k - 1\right)(\omega - 1)\omega\rho \le e^{\frac{k}{\omega-1}} \cdot \omega^2 \cdot \rho$$

for

$$\alpha = 1 + \frac{1}{\left(1 + \frac{1}{\omega-1}\right)^k - 1} \ge 1 + e^{\frac{-k}{\omega-1}} > 1.$$

Since Rényi divergence is an increasing function of the order α, this implies

$$\forall x, x' \in \mathcal{X}^n \; \mathrm{D}_1\left(M(x) \middle\| M(x')\right) \le e^{\frac{k}{\omega-1}} \cdot \omega^2 \cdot \rho.$$

Now, by the definition of mutual information and the convexity of KL divergence,

$$\begin{aligned} I(X; M(X)) &= \mathop{\mathbb{E}}_{x \leftarrow X}\left[\mathrm{D}_1\left(M(x) \| M(X)\right)\right] \\ &\le \mathop{\mathbb{E}}_{x, x' \leftarrow X}\left[\mathrm{D}_1\left(M(x) \middle\| M(x')\right)\right] \le e^{\frac{n}{\omega-1}} \omega^2 \rho. \end{aligned}$$

□

As an example, we prove a lower bound for point queries (cf. Corollary 18)

Corollary 27. *Let $M : \mathcal{X}^n \to \mathbb{R}^{\mathcal{X}}$ be (ρ, ω)-tCDP. Suppose*

$$\forall x \in \mathcal{X}^n \; \mathop{\mathbb{P}}_{M}\left[\max_{u \in \mathcal{X}} |M(x)_u - q_u(x)| \ge \frac{n}{2}\right] \le \frac{1}{2},$$

where $q_u(x) = |\{i \in [n] : x_i = u\}|$. Then

$$n \ge (\omega - 1) \log\left(\frac{\log(|\mathcal{X}|/4)}{2\omega^2 \rho}\right).$$

In contrast, our corresponding upper bound (Corollary 18) shows that such a M exists provided $n \ge 16\omega \operatorname{arsinh}(\sqrt{\log(2|\mathcal{X}|)/2\omega^2\rho})$. (Note $\operatorname{arsinh}(x) \approx \log(x)$.) That is, our lower bound is tight to within constant factors.

Proof. Let $U \in \mathcal{X}$ be uniformly random and let $X \in \mathcal{X}^n$ be n copies of U. By Theorem 26, $I(X; M(X)) \le e^{\frac{n}{\omega-1}} \omega^2 \rho$. Let $V = \operatorname{argmax}_{v \in \mathcal{X}} M(X)_v$. Since $q_U(X) = n$ and $q_{\widehat{U}}(X) = 0$ for any $\widehat{U} \ne U$, by our accuracy assumption, $\mathbb{P}[U = V] \ge 1/2$. Thus, by Fano's inequality,

$$\begin{aligned} e^{\frac{n}{\omega-1}} \omega^2 \rho &\ge I(X; M(X)) \ge I(U; V) \\ &= H(U) - H(U|V) \ge \frac{1}{2} \log\left(\frac{|\mathcal{X}|}{4}\right). \end{aligned}$$

Solving for n yields the result. □

ACKNOWLEDGMENTS

We thank the Isaac Newton Institute for Mathematical Sciences, Cambridge U.K., for support and hospitality during the programme "Data Linkage and Anonymisation" where part of this work was done under the support of EPSRC grant no EP/K032208/1.

G.N.R.'s research supported by the ISRAEL SCIENCE FOUNDATION (grant No. 5219/17).

REFERENCES

[1] Martín Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. 2016. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*.

[2] Mitali Bafna and Jonathan Ullman. 2017. The Price of Selection in Differential Privacy. In *Proceedings of the 30th Conference on Learning Theory, COLT 2017, Amsterdam, The Netherlands, 7-10 July 2017*. 151–168.

[3] Amos Beimel, Kobbi Nissim, and Uri Stemmer. 2013. Private Learning and Sanitization: Pure vs. Approximate Differential Privacy. In *APPROX-RANDOM*.

[4] Raghav Bhaskar, Srivatsan Laxman, Adam D. Smith, and Abhradeep Thakurta. 2010. Discovering frequent patterns in sensitive data. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, July 25-28, 2010*. 503–512.

[5] Mark Bun, Kobbi Nissim, Uri Stemmer, and Salil Vadhan. 2015. Differentially private release and learning of threshold functions. In *FOCS*. IEEE.

[6] Mark Bun and Thomas Steinke. 2016. Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds. In *TCC*. https://arxiv.org/abs/1605.02065.

[7] Kamalika Chaudhuri, Daniel J. Hsu, and Shuang Song. 2014. The Large Margin Mechanism for Differentially Private Maximization. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*. 1287–1295.

[8] Anindya De. 2012. Lower Bounds in Differential Privacy. In *TCC*. https://doi.org/10.1007/978-3-642-28914-9_18

[9] Cynthia Dwork. 2006. Differential Privacy. In *Proceedings of the 33rd International Conference on Automata, Languages and Programming - Volume Part II (ICALP'06)*. Springer-Verlag, Berlin, Heidelberg, 1–12. https://doi.org/10.1007/11787006_1

[10] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. 2006. Our Data, Ourselves: Privacy Via Distributed Noise Generation. In *EUROCRYPT*.

[11] Cynthia Dwork and Jing Lei. 2009. Differential privacy and robust statistics. In *STOC*.

[12] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In *TCC*. http://repository.cmu.edu/jpc/vol7/iss3/2.

[13] Cynthia Dwork, Moni Naor, Omer Reingold, and Guy N. Rothblum. 2015. Pure Differential Privacy for Rectangle Queries via Private Partitions. In *Advances in Cryptology - ASIACRYPT 2015 - 21st International Conference on the Theory and Application of Cryptology and Information Security, Auckland, New Zealand, November 29 - December 3, 2015, Proceedings, Part II*. 735–751.

[14] Cynthia Dwork and Guy Rothblum. 2016. Concentrated Differential Privacy. *CoRR* abs/1603.01887 (2016). http://arxiv.org/abs/1603.01887 https://arxiv.org/abs/1603.01887.

[15] Cynthia Dwork, Guy N. Rothblum, and Salil P. Vadhan. 2010. Boosting and Differential Privacy. In *FOCS*.

[16] Moritz Hardt and Kunal Talwar. 2010. On the Geometry of Differential Privacy. In *STOC*. https://doi.org/10.1145/1806689.1806786

[17] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam D. Smith. 2011. What Can We Learn Privately? *SIAM J. Comput.* 40, 3 (2011), 793–826.

[18] F. McSherry and K. Talwar. 2007. Mechanism Design via Differential Privacy. In *FOCS*. https://doi.org/10.1109/FOCS.2007.66

[19] Ilya Mironov. 2017. Renyi differential privacy. *arXiv preprint arXiv:1702.07476* (2017). https://arxiv.org/abs/1702.07476.

[20] Jack Murtagh and Salil P. Vadhan. 2016. The Complexity of Computing the Optimal Composition of Differential Privacy. In *TCC*.

[21] Alfréd Rényi. 1961. On Measures of Entropy and Information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. University of California Press, Berkeley, Calif., 547–561. http://projecteuclid.org/euclid.bsmsp/1200512181

[22] Adam Smith. 2009. Differential privacy and the secrecy of the sample. (2009). https://adamdsmith.wordpress.com/2009/09/02/sample-secrecy/.

[23] Thomas Steinke and Jonathan Ullman. 2017. Tight Lower Bounds for Differentially Private Selection. In *FOCS*. https://arxiv.org/abs/1704.03024.

[24] T. van Erven and P. Harremos. 2014. Rényi Divergence and Kullback-Leibler Divergence. *IEEE Transactions on Information Theory* 60, 7 (July 2014), 3797–3820. https://doi.org/10.1109/TIT.2014.2320500 https://arxiv.org/abs/1206.2459.