Additive Noise Mechanisms for Making Randomized Approximation Algorithms Differentially Private

Jakub Tětek
j.tetek@gmail.com
Basic Algorithms Research Copenhagen,
University of Copenhagen

Abstract

The exponential increase in the amount of available data makes taking advantage of them without violating users’ privacy one of the fundamental problems of computer science. This question has been investigated thoroughly under the framework of differential privacy. However, most of the literature has not focused on settings where the amount of data is so large that we are not even able to compute the exact answer in the non-private setting (such as in the streaming setting, sublinear-time setting, etc.). This can often make the use of differential privacy unfeasible in practice.

In this paper, we show a general approach for making Monte-Carlo randomized approximation algorithms differentially private. We only need to assume the error $R$ of the approximation algorithm is sufficiently concentrated around 0 (e.g. $E[|R|]$ is bounded) and that the function being approximated has a small global sensitivity $\Delta$. Specifically, if we have a randomized approximation algorithm with sufficiently concentrated error which has time/space/query complexity $T(n, \rho)$ with $\rho$ being an accuracy parameter, we can generally speaking get an algorithm with the same accuracy and complexity $T(n, \Theta(\epsilon \rho))$ that is $\epsilon$-differentially private.

Our technical results are as follows. First, we show that if the error is subexponential, then the Laplace mechanism with error magnitude proportional to the sum of the global sensitivity $\Delta$ and the subexponential diameter of the error of the algorithm makes the algorithm differentially private. This is true even if the worst-case global sensitivity of the algorithm is large or infinite. We then introduce a new additive noise mechanism, which we call the zero-symmetric Pareto mechanism. We show that using this mechanism, we can make an algorithm differentially private even if we only assume a bound on the first absolute moment of the error $E[|R|]$.

Finally, we use our results to give either the first known or improved sublinear-complexity differentially private algorithms for various problems. This includes results for frequency moments, estimating the average degree of a graph in sublinear time, rank queries, or estimating the size of the maximum matching. Our results raise many new questions and we state multiple open problems.
1 Introduction

With the increase in the amount of available data, the problem of analyzing it in a privacy-preserving manner has become a central problem in computer science. One commonly used tool for this task is differential privacy, which is a well-established notion of privacy that is commonly used in data analysis and machine learning. However, with some notable exceptions, the literature on differential privacy has focused on the setting where the amount of data is small enough that we would be able to practically solve a given problem exactly in the non-private setting. However, in practice, this assumption is often not realistic – this is after all the reason for the existence of (among others) streaming and sublinear-time algorithms.

Our goal is thus to get very efficient (sublinear) algorithms that at the same time guarantee differential privacy. In the streaming or sublinear-time setting, the error coming from the algorithm not being exact will generally speaking be much bigger than the amount of noise that the given problem necessitates for ensuring differential privacy. The main objective in this setting is thus not to simply minimize the amount of noise we add, but rather to achieve a given level of accuracy while minimizing the complexity (e.g. space, time, or query complexity) of the algorithm. Of course, some amount of noise inherently has to be added to achieve privacy, but this is usually so small, that one would need linear complexity to get such a level of accuracy even without privacy. In the sublinear regime, we thus usually do not have to worry whether a given level of accuracy is achievable and we instead focus as our central objective on the complexity needed to achieve it.

One of the main difficulties in making sublinear algorithms private is that most sublinear-time and streaming algorithms are randomized and give only probabilistic guarantees on the quality of the output. This makes adding noise based on global sensitivity\(^1\) – which is commonly used to get differentially private algorithms – unsuitable for this situation, as in the worst case the global sensitivity of the approximation algorithm\(^2\) can be very large even if the global sensitivity of the function being approximated is small. In this paper, we propose a way to get around this issue by showing additive noise mechanisms that only need that (1) the function being approximated has low global sensitivity and (2) the answer of the algorithm is sufficiently concentrated around the true value.

We give two main results – one under the assumption that the error has subexponential tails, while the other only assumes bounded mean deviation (or higher moments). While the first one has a much stronger assumption about the error distribution, it is stronger in that it also works for multiple adaptive queries that are not answered independently. This is useful for example for streaming algorithms, where multiple queries can be answered using a single sketch and are thus not answered independently. Note that the standard composition theorem would allow us to perform multiple queries only if they were answered independently.

We use our results to give new differentially private algorithms for various problems: for maximum matching under node-level privacy, frequency moments, counting connected components under edge-level privacy, and rank queries. We also show how a common technique for designing relative-approximation sublinear-time algorithms – advice removal by geometric search – can be made differentially private. This implies an edge-differentially-private algorithm for estimating the average degree of a graph, improving upon the state of the art \([6]\), but we think it could also be useful for many other problems. Our algorithm for maximum matching also answers an open problem from \([6]\). For the other mentioned problems, we either give the first sublinear-complexity algorithm that ensures pure differential privacy, or one that is more efficient.

---

1 Global sensitivity of a function \( g \) with respect to a relation \( \sim \) is defined as \( \sup_{x \sim x'} |g(x) - g(x')| \).
2 Here, we see the approximation algorithm as a deterministic function of the input and a string of random bits.
11.1 High-level view of technical results

We now informally state our main technical results and sketch how we use them in order to get private algorithms.

Subexponential error tails. Our first central result is as follows:

**Theorem 10** simplified version. Assume we have an algorithm \(A(D)\) for \(D\) being a dataset. Assume there exists a function \(g\) with global sensitivity \(\leq \Delta_1\) w.r.t. \(D\) such that \(A(D) - g(D)\) has subexponential diameter \(\leq \Delta_2\), i.e. \(\Pr[|A(D) - g(D)| \geq t] \leq 2e^{-t/\Delta_2}\) for \(t \geq 0\). Then releasing \(A(D) + \text{Laplace}(O((\Delta_1 + \Delta_2)/\epsilon))\) is \(\epsilon\)-differentially private for \(\epsilon \leq O(1)\).

Moreover, with noise \(\text{Laplace}(O(k(\Delta_1 + \Delta_2)/\epsilon))\), this also holds if we make \(k\) such releases with different algorithms \(A_1, \ldots, A_k\) chosen adaptively that are executed with the same randomness.

Note that this generalizes the claim that the Laplace mechanism with noise magnitude proportional to the global sensitivity gives differential privacy (this can be seen by setting \(A(D) = g(D)\)). Note also that in the second half, the algorithms use the same randomness, and we thus cannot get this part of the result by standard composition.

One would usually use this theorem as follows. We start with a randomized approximation algorithm whose error is subexponentially concentrated around zero (often, this is either known or easy to prove) that approximates a parameter with a small global sensitivity \(\Delta_1\). This is the case for example for the YYI maximum matching algorithm [36] under node-level privacy or the KLL sketch [21] for rank queries. Suppose the complexity (such as time/space/query/sample or other complexity) of the algorithm is \(T(n, \rho)\) and has additive error of scale (i.e. with subexponential diameter) \(\rho n\). If we want the final error with privacy to be \(O(\rho n)\), then we run the algorithm with error parameter \(\epsilon \rho\), making the error’s subexponential diameter be \(O(\epsilon \rho n)\). We then get from Theorem 10 that adding noise of magnitude \(O(\rho n)\) is sufficient to get \(\epsilon\)-differential privacy, assuming \(\Delta_1\) is sufficiently small, giving us the desired result. The complexity of the private algorithm will thus be \(T(n, \Theta(\epsilon \rho))\). This allows us to achieve a given level of accuracy under pure differential privacy, while not significantly worsening the algorithm’s complexity. We describe this approach in greater detail and with general failure probabilities (not just constant) in Section 5.

To illustrate the second part of the theorem, consider for example a rank queries sketch (see Section 5.5 for details). The algorithms \(A_1, \ldots, A_k\) correspond to making \(k\) adaptive queries to the sketch of a dataset (assume we are given \(k\) queries we need to perform) and the fact that the algorithms use the same randomness corresponds to us querying the same sketch (as compared to \(k\) independent sketches). Specifically, the algorithm \(A_i\) here builds the sketch (using the shared randomness), and then performs the \(i\)-th query. Note that the fact that we only use one sketch prevents us from using the composition theorem to get this from just the first part of the theorem.

**Polynomial error tails.** We then prove that in the case of a single query, it is sufficient to assume a bound on some deviation moments (multiple independently answered queries can be handled using the standard composition theorem).

**Theorem 14** simplified version. Let us have an algorithm \(A\) such that there exists a function \(g\) with global sensitivity \(\leq \Delta\) and such that \(\mathbb{E}[|A(D) - g(D)|^3] \leq \Delta^3\) for any dataset \(D\). Then for \(\epsilon \leq O(1)\) there exists a random variable \(Y\) with \(\mathbb{E}[|Y|] \leq O(\Delta/\epsilon)\), such that \(A(\cdot) + Y\) is \(\epsilon\)-differentially private.

This theorem can be used in a way similar to Theorem 10 whose use we described above. Moreover, it holds that if we have \(\mathbb{E}[|A(D) - g(D)|] \leq \Delta\), then by taking a median of 5 independent executions of

\[\begin{equation}
\text{This is true unless } \rho \text{ is very small, as otherwise the global sensitivity will necessitate some level of noise. As we noted, this usually happens only when } T(n, \rho) \geq O(n), \text{ making this uninteresting for our sublinear setting.}
\end{equation}\]
Assume for simplicity that both $\Delta$ have subexponential diameter $\Delta_2 \leq 1$. The same approach works for general values of $\Delta_1, \Delta_2$ by simple re-scaling. Let $Y \sim \text{Laplace}(c/\epsilon)$ for appropriately chosen value of $c$. We will prove that for any random variable $X$ with subexponential diameter $\leq 3$ (that is $\mathbb{P}[|X| \geq t] \leq 2e^{-t/3}$), the probability density functions satisfy $f_{X+Y}(y)/f_Y(y) = e^{\pm \Theta(c)}$ for any $y$. We can use this to prove privacy, as we now show. For two neighboring datasets $D_1, D_2$, we set $R_1 = A(D_1) - g(D_1)$ and $R_2 = A(D_2) - g(D_1)$ (note the asymmetry in the definitions). It then holds that $R_1$ has subexponential diameter $\Delta_2 \leq 1$. One can also show that the subexponential diameter $\Delta_2$ can be decreased at a rate of $O(1)$ (since $\Delta_2$ is an upper bound on the “subexponential diameter” of $R$). Intuitively speaking, $\Delta_2$ determines the “scale” of $R$, and $\Delta_2$ is in fact up to a constant factor an upper bound on $\mathbb{E}(|R|)$. Note that the tails of $R$ decrease at least at the same rate as those of the Laplace distribution. This suggests Laplacian noise with large enough magnitude could “hide $R$”. Indeed, we prove that Laplacian noise will guarantee privacy. We now sketch the proof.

**High-level view.** Assume for simplicity that both $\Delta_1, \Delta_2 \leq 1$. The same approach works for general values of $\Delta_1, \Delta_2$ by simple re-scaling. Let $Y \sim \text{Laplace}(c/\epsilon)$ for appropriately chosen value of $c$. We will prove that for any random variable $X$ with subexponential diameter $\leq 3$ (that is $\mathbb{P}[|X| \geq t] \leq 2e^{-t/3}$), the probability density functions satisfy $f_{X+Y}(y)/f_Y(y) = e^{\pm \Theta(c)}$ for any $y$. We can use this to prove privacy, as we now show. For two neighboring datasets $D_1, D_2$, we set $R_1 = A(D_1) - g(D_1)$ and $R_2 = A(D_2) - g(D_1)$ (note the asymmetry in the definitions). It then holds that $R_1$ has subexponential diameter $\Delta_2 \leq 1$. One can also show that the subexponential diameter $\Delta_2$ can be decreased at a rate of $O(1)$ (since $\Delta_2$ is an upper bound on the “subexponential diameter” of $R$). Intuitively speaking, $\Delta_2$ determines the “scale” of $R$, and $\Delta_2$ is in fact up to a constant factor an upper bound on $\mathbb{E}(|R|)$. Note that the tails of $R$ decrease at least at the same rate as those of the Laplace distribution. This suggests Laplacian noise with large enough magnitude could “hide $R$”. Indeed, we prove that Laplacian noise will guarantee privacy. We now sketch the proof.

**Subexponential error, one query.** Suppose we have a randomized algorithm $A(D)$ for $D$ being a dataset that approximates a function $g(D)$ with global sensitivity $\leq \Delta_1$ for some parameter $\Delta_1$. Define the error $R$ as the random variable $R = A(D) - g(D)$ and assume that it is tightly concentrated around 0, namely $\mathbb{P}[|R| > t] \leq 2e^{-t/\Delta_2}$ for some value $\Delta_2$ (where $\Delta_2$ is an upper bound on the “subexponential diameter” of $R$). Intuitively speaking, $\Delta_2$ determines the “scale” of $R$, and $\Delta_2$ is in fact up to a constant factor an upper bound on $\mathbb{E}(|R|)$. Note that the tails of $R$ decrease at least at the same rate as those of the Laplace distribution. This suggests Laplacian noise with large enough magnitude could “hide $R$”. Indeed, we prove that Laplacian noise will guarantee privacy. We now sketch the proof.

**High-level view.** Assume for simplicity that both $\Delta_1, \Delta_2 \leq 1$. The same approach works for general values of $\Delta_1, \Delta_2$ by simple re-scaling. Let $Y \sim \text{Laplace}(c/\epsilon)$ for appropriately chosen value of $c$. We will prove that for any random variable $X$ with subexponential diameter $\leq 3$ (that is $\mathbb{P}[|X| \geq t] \leq 2e^{-t/3}$), the probability density functions satisfy $f_{X+Y}(y)/f_Y(y) = e^{\pm \Theta(c)}$ for any $y$. We can use this to prove privacy, as we now show. For two neighboring datasets $D_1, D_2$, we set $R_1 = A(D_1) - g(D_1)$ and $R_2 = A(D_2) - g(D_1)$ (note the asymmetry in the definitions). It then holds that $R_1$ has subexponential diameter $\Delta_2 \leq 1$. One can also show that the subexponential diameter $\Delta_2$ can be decreased at a rate of $O(1)$ (since $\Delta_2$ is an upper bound on the “subexponential diameter” of $R$). Intuitively speaking, $\Delta_2$ determines the “scale” of $R$, and $\Delta_2$ is in fact up to a constant factor an upper bound on $\mathbb{E}(|R|)$. Note that the tails of $R$ decrease at least at the same rate as those of the Laplace distribution. This suggests Laplacian noise with large enough magnitude could “hide $R$”. Indeed, we prove that Laplacian noise will guarantee privacy. We now sketch the proof.

---

4They also consider functions with low smoothed sensitivity instead of just low global sensitivity; we do not consider that in this paper.

5We choose value 3 as this is the value we will need below. The claim holds also for larger constants with $c$ chosen appropriately.
diameter of $R_2 = (A(D_2) - g(D_2)) + (g(D_2) - g(D_1))$ is $\leq 3$ (Lemma [6]). Let $y' = y - g(D_1)$. It then holds for any $y$ that

$$\frac{f_{A(D_1)+Y}(y)}{f_{A(D_2)+Y}(y)} = \frac{f_{g(D_1)+R_1+Y}(y)}{f_{g(D_1)+R_2+Y}(y)} = \frac{f_{R_1+Y}(y')}{f_{R_2+Y}(y')} = \frac{f_{Y}(y')}{f_{Y}(y')} = e^{\pm \Theta(\epsilon)}$$

where the last equality uses the claim $f_{X+Y}(y')/f_Y(y') = e^{\pm \Theta(\epsilon)}$ for $X = R_1$ and for $X = R_2$. This implies differential privacy.

Bounding ratios of density functions. We now sketch why $f_{X+Y}(y)/f_Y(y) = e^{\pm \Theta(\epsilon)}$. Since $Y$ is continuous, we may re-write

$$f_{X+Y}(y) = \mathbb{E}[f_Y(y - X)] = \frac{e}{2c} \mathbb{E}[\exp(-|X - y|/c)] = (*)$$

where the first equality is a standard identity [17]. We now use the inequalities $|X - y| \leq |X| + |y|$ and $|X - y| \geq |y| - |X|$. This allows to bound

$$(*) \leq \frac{e}{2c} \mathbb{E}[\exp(-\epsilon(|y| - |X|)/c)] = \frac{e}{2c} e^{-\epsilon|y|/c} \mathbb{E}[\exp(\epsilon|X|/c)]$$

$$(*) \geq \frac{e}{2c} \mathbb{E}[\exp(-\epsilon(|X| + |y|)/c)] = \frac{e}{2c} e^{-\epsilon|y|/c} \mathbb{E}[\exp(-\epsilon|X|/c)]$$

while it holds that $f_Y(y) = \frac{e}{2c} e^{-\epsilon|y|/c}$. It is thus sufficient to prove that $\mathbb{E}[\exp(\epsilon|X|/c)] \leq e^{O(\epsilon)}$ and $\mathbb{E}[\exp(-\epsilon|X|/c)] \geq e^{-O(\epsilon)}$. While the first inequality is standard, the second is not. We will now sketch a proof for both.

Bounding the expectation of exponentials of a subexponential random variable. If we knew the density function of $X$, we could easily express the expectations as integrals. However, not only we do not have a bound on the density, but the density may even not exist. We thus use the following trick. Unlike the density function, we do have a bound on the cumulative distribution function. Specifically, we are assuming $\mathbb{P}[|X| > \ell] \leq 2e^{-\ell/3}$ which implies that $F_{\epsilon|X|/c}^{-1}(u) \leq -\frac{3c}{2\epsilon} \log(\frac{1}{2u})$. Upper-bounding the CDF like this reduces the problem to computing the expectation of a function of the uniform random variable, which can be done straightforwardly. This proves the desired bounds.

Subexponential error, multiple queries. We would like to be able to release answers to multiple queries which are not answered independently (such as if they are answered based on the same sketch). Since the answers are not independent, we cannot use the composition theorem. We now sketch an alternative approach.

For fixed queries, the above proof goes through with minor modifications even in the multivariate case. Instead of using the inequalities $|y - X| \leq |y| + |X|$ and $|y - X| \geq |y| - |X|$ in the case of $y, X \in \mathbb{R}$, we use the analogous version for $\ell_1$ norms in the case of $y, X \in \mathbb{R}^k$: $\|y - X\| \leq \|y\| + \|X\|$ and $\|y - X\| \geq \|y\| - \|X\|$. We then use that if $X$ is a vector of $k$ subexponential random variables with diameter $\leq \Delta$, then $\|X\|_1$ has subexponential diameter $\leq 3k\Delta$ (Lemma [6]).

This however gives the result only in the nonadaptive case, when the queries do not depend on the released values: the identity $f_{X+Y}(y) = \mathbb{E}[f_Y(X - y)]$ relies crucially on $X = (X_1, \ldots, X_k)$ and $Y = (Y_1, \ldots, Y_k)$ being independent. In the case of adaptive queries, $X_i$ could depend on
Y_1, \cdots, Y_{i-1} (the query that we perform – and thus also the answer to it – can be influenced by the
noise we add to the previous answers). We instead use our non-adaptive version of the claim and
make it adaptive in a black box fashion, by proving a claim that may be of independent interest:
If we have a countable number of mechanisms such that releasing the answers of any fixed
subset of size k is \( \epsilon \)-differentially private, then we may also pick this subset adaptively and it will still be
\( \epsilon \)-differentially private.

**Error with polynomial tails.** In the case that the error has polynomial tails, we only consider
the case of a single query. Our techniques do not seem to generalize to the multivariate case, and
we conjecture that this is impossible (see Section 6). The case when multiple queries are answered
independently may be still handled by the standard composition theorem.

An approach similar to the one described above can be made to work, with the difference that
we use the inequality \( |x - X| \geq \max(0, |x| - |X|) \) instead of the weaker \( |x - X| \geq |x| - |X| \). This
approach, however, requires proving the following inequality for all \( y, s \geq 0, \alpha > 1, 0 \leq \epsilon \leq 1: \)

\[
\int_0^1 \min \left( (1 + |y|/s)\alpha, 1 - \frac{(1 - 2^{-1/\alpha})\epsilon}{(1 + |y|/s)(1 - u)^{1/\alpha}} \right)^{-\alpha} du \leq 1 + \frac{2\alpha - 1}{\alpha - 1} \epsilon.
\]

This is the technically most challenging part of this paper. The trick is to bound the inside of the
integral for \( u \in [0, 1 - \epsilon(1 + |y|/s)^{-\alpha}] \) by a simpler expression that can be successfully integrated.
The rest of the interval \([0, 1]\) contributes at most \( \epsilon \), as its length is \( \epsilon(1 + |y|/s)^{-\alpha} \) and the maximum
value of the function being integrated is \( \leq (1 + |y|/s)^\alpha \).

1.3 Related work

To the best of our knowledge, the work on differentially private approximation algorithms started
with private sketches. Mir, Muthukrishnan, Nikolov, and Wright [26] gave pan-private sketches for
heavy hitters. An improved sketch has been recently given by Pagh and Thorup [28]. A private
version of the deterministic Misra-Gries sketch [27] for heavy hitters has been recently given by
Lebeda and Tétek [23]. Heavy hitters were also investigated in the multi-party computation setting [19],
in the local differential privacy setting [4], and using cryptographic assumptions [25, 18, 19].

A sketch for fractional frequency moments \( F_p \) for \( 0 \leq p \leq 1 \) has been given by Wang, Pinelis, and
Song [35]. After releasing this paper, Epasto, Mao, Medina, Mirrokni, Vassilvitskii, and Zhong [16]
gave an algorithm general value of \( p \) in the continual release setting. A sketch for differentially
private quantiles has been given by Alabi, Ben-Eliezer, and Chaturvedi [1]. A technique for stream
sanitization has been given by Kaplan and Stemmer [20]; this work resulted in improved differentially
private sketches for approximate quantiles. An approach for differentially privately estimating
distances in euclidean spaces using private sketches has been given by Stausholm [33]. A general
approach to making linear sketches differentially private was given by Zhao, Qiao, Redberg, Agrawal,
Abbadi, and Wang [37].

A recent line of work has shown that many sketches already provide privacy by themeselves or
with only small modifications, without adding any noise. Blocki, Blum, Datta, and Sheffet [7] have
shown that the Johnson-Lindenstrauss transform by itself ensures differential privacy. Smith, Song,
and Guha Thakurta [32] have shown that this is also the case for the Flajolet-Martin sketch for
counting distinct elements and Choi, Dachman-Soled, Kulkarni, and Yerukhimovich [10] have given

---

\[\text{An algorithm on an input stream is said to be pan-private if releasing the internal state of the algorithm at any}\]
\[\text{point in the computation is differentially private. It is a strictly stronger notation than differential private of the}\]
\[\text{output.} \]
a similar result for the LogLog algorithm [12]. This was recently generalized by Dickens, Thaler, and Ting [11] who show that this is not only the case for the two above-mentioned sketches, but in fact for a large class of sketches for counting distinct elements.

As far as sublinear-time algorithms are concerned, Sivasubramaniam, Li, and He [31] have shown a differentially private algorithm that returns a $2 + \rho$ approximation of the number of edges in a graph in time $\tilde{O}(\rho, \sqrt{n})$. This has been later improved by Blocki et al. [6] to $1 + \rho$ approximation in the same complexity. In that paper, the authors also give differentially private sublinear algorithms for approximate maximum matching and vertex cover. A sublinear time algorithm for estimating the median was recently given by Boehler and Kerschbaum [9].

2 Preliminaries

2.1 Differential privacy

Throughout the paper, we assume that we have a symmetric “neighbor” relation $\sim$ on the set of all possible datasets. Intuitively speaking, in the case when we have a database of users, this should correspond to two datasets being the same except for the data of one user whose privacy we are trying to protect.

Definition 1 [13]. A randomized algorithm $M$ with range $S$ is $\epsilon$-differentially private if for any measurable $T \subseteq S$, it holds for any $x \sim x'$ for a symmetric “neighbor” relation $\sim$, that

$$e^{-\epsilon} \leq \frac{\Pr[M(x) \in T]}{\Pr[M(x') \in T]} \leq e^\epsilon$$

This definition is commonly relaxed to a notion called approximate differential privacy, with the above notion then being called pure privacy. In this paper, we will focus only on pure differential privacy.

If the output of $M$ is a continuous random variable, then it is sufficient to prove that for any $y$ and $x \sim x'$ it holds $e^{-\epsilon} \leq f_{M(x)}(y)/f_{M(x')}(y) \leq e^\epsilon$, where $f_X$ for $X$ being a continuous random variable is the probability density function of $X$.

The global sensitivity [14] of a function $g$ is defined as

$$\sup_{x \sim x'} |g(x) - g(x')|.$$ 

Dwork et al. [14] have shown that if $g$ has global sensitivity $\Delta$, then adding Laplace($\Delta/\epsilon$) provides $\epsilon$-differential privacy.

In the context of graph problems, one often speaks of a mechanism being edge-differentially private, or node-differentially private. These terms refer to the relation $\sim$ that is used. In the case of node-differential privacy, we have $G \sim G'$ iff one can get $G$ from $G'$ by deleting one vertex and the incident edges, or the other way around. In the case of edge-differential privacy, we have $G \sim G'$ iff one can get $G$ from $G'$ by deleting one edge, or the other way around.

2.2 Probability theory

If $D$ is a distribution, we use $D^\otimes k$ for $k$ being a natural number, to denote the $k$-fold product distribution of $D$. For a random variable $Z$, we denote by $F_Z$ its cumulative distribution function. We denote by $F_Z^{-1}(p) = \inf\{x \in \mathbb{R} : F_Z(x) \geq p\}$ its generalized inverse. It holds that $F_Z^{-1}(u)$ has the same distribution as $Z$ for $u \sim \text{Unif}(0, 1)$ [24]. We will need the following claim.
Fact 2 (\[17\]). Let $X, Y$ be independent random variables in $\mathbb{R}^k$, and assume $Y$ has a probability density function (pdf) $f_Y(z)$. Then the pdf of $X + Y$ is $f_{X+Y}(z) = \mathbb{E}[f_Y(z - X)]$.

We will also need the following claim:

Lemma 3 (Moment amplification, \[22\]). Let $X_1, \cdots, X_{2k-1}$ be i.i.d. random variables on $\mathbb{R}$. It holds for any $x, c \in \mathbb{R}$ that

$$E[|\text{median}(X_1, \cdots, X_{2k-1}) - x|^{ck}]^{1/ck} \leq O(E[|X_1 - x|^{c}]^{1/c})$$

Concentration of measure. The notions of subexponential random variables and subexponential diameter $\sigma_{se}[X]$ are central to concentration of measure. There are several different definitions for $\sigma_{se}[X]$, that differ by constant factors. See for example \[34, Chapter 2\] for an exposition of the various definitions. In this paper, one of the definitions is especially suitable for the way we use it in our proofs, and that is the definition that we use.

Definition 4. Let $X$ be a real random variable. We define the subexponential diameter of $X$, denoted by $\sigma_{se}[X]$ as the smallest values for which for any $t > 0$ holds

$$\mathbb{P}[|X| \geq t] \leq 2 \exp(-t/\sigma_{se}[X])$$

A random variable $X$ is said to be subexponential if $\sigma_{se}[X] < \infty$.

It holds that $\sigma_{se}[cX] = c\sigma_{se}[X]$. It also holds that

Fact 5. For $X$ being a random variable and $c \geq 0$, it holds that $\sigma_{se}[X + c] \leq \sigma_{se}[X] + c/ \log 2$.

Proof. We can bound $\mathbb{P}[X + c \geq t] \leq \min(1, 2e^{-(t-c)/\sigma_{se}[X]}) \leq \min(1, 2e^{-t/(\sigma_{se}[X]+c/\log 2)})$, thus implying the claim.

Finally, the following is a standard claim, but we need the constant factor, which is specific to the definition of $\sigma_{se}$ that we are using. We thus give a proof in Appendix A based on the standard proof of triangle inequality for Orlicz norms.

Lemma 6. Let us have real random variables $X_1, \cdots, X_k$. It holds

$$\sigma_{se}\left[\sum_{i=1}^{k} X_i\right] \leq 3 \sum_{i=1}^{k} \sigma_{se}[X_i]$$

3 Algorithms with subexponential error

In this section, we show how algorithms, whose error has a small subexponential diameter, can be made differentially private. We start by proving a technical lemma. We will later bound the ratios between probability densities of our mechanism’s answers by the exponential expectations that we now bound and, finally, we will use that to prove privacy in Theorem 10, which is the main theorem of this section.

Note that while the second of the two inequalities is standard, the first one is not. Our proof does not follow the strategy of the standard proof of the second inequality, which is based on a Taylor expansion and does not seem to straightforwardly apply to the first inequality.

Lemma 7. Suppose $X$ is a random variable with subexponential diameter $\Delta \leq 1/2$. It holds

$$\mathbb{E}[e^{-|X|}] \geq 2^{-\Delta} \geq e^{-(1+\log 2)\Delta} \text{ and } \mathbb{E}[e^{|X|}] \leq 2^{\Delta} \leq e^{3\log(2)\Delta}.$$
Proof. Since $X$ is subexponential with diameter $\Delta$, it holds that $\mathbb{P}[|X| \geq z] \leq 2e^{-z/\Delta}$. Therefore, $F^{-1}_{|X|}(u) \leq -\Delta \log(\frac{1-u}{2})$. We use the fact that for $u \sim Unif(0,1)$, the random variable $F^{-1}_{|X|}(u)$ has the same distribution as $|X|$. We can now bound

$$
\mathbb{E}[e^{-|X|}] = \mathbb{E}_u[e^{-F^{-1}_{|X|}(u)}] \\
\geq \mathbb{E}_u[e^{\Delta \log(\frac{1-u}{2})}] \\
= 2^{-\Delta} \mathbb{E}_u[(1-u)^{\Delta}] \\
= 2^{-\Delta} \int_0^1 (1-u)^{\Delta} du \\
= 2^{-\Delta} \left[ \frac{(1-u)^{\Delta+1}}{\Delta+1} \right]_{u=0}^{1} \\
= \frac{2^{-\Delta}}{1+\Delta} \geq e^{-(1+\log 2)\Delta}
$$

where the last inequality holds because we can equivalently write $2^{-\Delta}/(\Delta + 1) \geq (2e)^{-(\Delta)}$, which simplifies to $e^\Delta \geq 1 + \Delta$, which is a standard inequality. Similarly, we can bound $\mathbb{E}[e^{|X|}]$ as follows.

$$
\mathbb{E}[e^{|X|}] = \mathbb{E}_u[e^{F^{-1}_{|X|}(u)}] \\
\leq \mathbb{E}_u[e^{-\Delta \log(\frac{1-u}{2})}] \\
= \frac{2^\Delta}{1-\Delta} \leq e^{3\log(2)\Delta}
$$

where the second equality is by substituting $\Delta$ with $-\Delta$ in (1) (since as we have shown, (1) is equal to $2^{-\Delta}/(\Delta + 1)$). We have here used that $\Delta < 1$ (otherwise the final expression may not be defined). The last inequality can be shown as follows: we take the ratio of the two sides resulting in $h(\Delta) = 4^\Delta (1-\Delta)$ and we show $h'(\Delta) \geq 1$ for $0 \leq \Delta \leq 1/2$. The function $h$ is concave (the second derivative is $-2 \cdot 2^\Delta \log 4 + 2^\Delta \log 2 \cdot 4$, which can be easily seen to be negative), meaning that it is sufficient to check that the inequality holds at the endpoints of the interval $[0, 1/2]$: that $h(0) \geq 1$ and $h(1/2) \geq 1$. One can easily check this holds. \qed

We are now ready to prove a lemma about the ratio of the density function of a Laplace and of Laplace shifted by a subexponential random variable. We will then use this to prove differential privacy. Note that the random variables in the lemma do not have to be independent.

Lemma 8. Let $X_1, \ldots, X_k$ be random variables with subexponential diameter at most $\Delta$ and let $X = (X_1, \ldots, X_k)$. Let $Y \sim \text{Lap}^k(\kappa \Delta/\epsilon)$ for $\epsilon \leq 1/6$. Consider $y \in \mathbb{R}^k$. It holds $e^{-3(1+\log 2)\kappa} \leq f_{X+Y}(y)/f_{Y}(y) \leq e^{3\log(2)\kappa}$. Moreover, if $k = 1$ and $\epsilon \leq 1/2$, it holds $e^{-3(1+\log 2)\kappa} \leq f_{X+Y}(y)/f_Y(x)$.

Proof. By Fact 2, we have

$$
f_{X+Y}(y) = \mathbb{E}[f_Y(y - X)] = \left( \frac{\epsilon}{2k\Delta} \right)^k \mathbb{E}[\exp(-\frac{\epsilon k\Delta}{X - y})]
$$

For the sake of brevity, we let $\gamma = \left( \frac{1}{2\epsilon k\Delta} \right)^k$. We may bound $\|X - y\|_1 \leq \|X\|_1 + \|y\|_1$ and $\|X - y\|_1 \geq \|y\|_1 - \|X\|_1$. This allows us to bound

$$
f_{X+Y}(y) = \gamma \mathbb{E}\left[ \exp\left( -\frac{\epsilon \|X - y\|_1}{k\Delta} \right) \right]
$$

8
\begin{align*}
\geq \gamma E \left[ \exp \left( \frac{-\epsilon (\|X \|_1 + \|y \|_1)}{k \Delta} \right) \right] \\
= \gamma \exp \left( \frac{-\epsilon \|y \|_1}{k \Delta} \right) E \left[ \exp \left( \frac{-\epsilon \|X \|_1}{k \Delta} \right) \right] \\
\geq \gamma \exp \left( \frac{-\epsilon \|y \|_1}{k \Delta} \right) \exp(-3(1 + \log 2)\epsilon)
\end{align*}

where the last inequality is by Lemma 7; we used that \( \epsilon \|X\|_1 \) has subexponential diameter \( \leq 3\epsilon \leq 1/2 \), since we have by Lemma 6 that \( \sigma_{se}[\|X\|_1] \leq 3k\Delta \). In the case \( k = 1 \), we simply have that \( \sigma_{se}[\|X\|_1] \leq \epsilon \), in which case the final bound on (2) is

\[ \geq \gamma \exp \left( \frac{-\epsilon \|y \|_1}{k \Delta} \right) \exp(-1 + \log 2)\epsilon \]

At the same time, \( f_Y(y) = \gamma \exp\left(\frac{-\epsilon \|y\|_1}{k \Delta}\right) \) and thus

\[ \frac{f_{X+Y}(y)}{f_Y(y)} \geq e^{-3(1+\log 2)\epsilon} \quad \text{if } k > 1 \]
\[ \frac{f_{X+Y}(y)}{f_Y(y)} \geq e^{-(1+\log 2)\epsilon} \quad \text{if } k = 1 \]

Similarly, we can bound

\[ f_{X+Y}(y) = \gamma E \left[ \exp \left( \frac{-\epsilon \|X - y \|_1}{k \Delta} \right) \right] \]
\[ \leq \gamma E \left[ \exp \left( \frac{-\epsilon (\|y \|_1 - \|X \|_1)}{k \Delta} \right) \right] \]
\[ = \gamma \exp \left( \frac{-\epsilon \|y \|_1}{k \Delta} \right) E \left[ \exp \left( \frac{\epsilon \|X \|_1}{k \Delta} \right) \right] \]
\[ \leq \gamma \exp \left( \frac{-\epsilon \|y \|_1}{k \Delta} \right) \exp(9 \log(2)\epsilon) \]

in the case \( k > 1 \) and

\[ f_{X+Y}(y) \leq \gamma \exp \left( \frac{-\epsilon \|y \|_1}{k \Delta} \right) \exp(3 \log(2)\epsilon) \]

in the case \( k = 1 \). Thus, we have

\[ \frac{f_{X+Y}(y)}{f_Y(y)} \leq e^{9\log(2)\epsilon} \quad \text{if } k > 1 \]
\[ \frac{f_{X+Y}(y)}{f_Y(y)} \leq e^{3\log(2)\epsilon} \quad \text{if } k = 1 \]

We now prove a lemma that says that in general, if we have a countable number of mechanisms and releasing any fixed \( k \) of them is differentially private, then picking the mechanisms adaptively will not violate differential privacy. The proof roughly follows the outline of the proof of adaptive composition \cite{29}. In what follows, we again use the subscript \( \cdot_{\cdot_{\cdot}} \) to denote a random bitstring used as the source of randomness of the mechanisms. NOTE TO REVIEWERS: The reviewers may wish to skip this proof and go to page 11 instead as the following proof technique is quite standard. On page 11, we finish the proof of the main theorem.
Lemma 9. Let us have a countable number of mechanisms $M_{1,r}, \ldots, M_{i_k,r}(D)$, such that releasing the value $(M_{i_1,r}(D), \ldots, M_{i_k,r}(D))$ is $\epsilon$-differentially private for any fixed $i_1, \ldots, i_k \in [n]$. Then the mechanism $(M_{j_1,r}(D), \ldots, M_{j_k,r}(D))$ is $\epsilon$-differentially private for $j_1, \ldots, j_k \in [n]$ such that $j_\ell$ for $\ell \in [k]$ is drawn from a distribution which is a function of $M_{j_1,r}(D), \ldots, M_{j_{k-1},r}(D)$.

Proof. Let $V$ be the adversary selecting the values of $j_\ell$. At the end, $V$ releases $M_{j_1,r}(D), \ldots, M_{j_k,r}(D)$; we prove $V$ is differentially private. At step $\ell$, $V$ picks a distribution as a function of $M_{j_1,r}(D), \ldots, M_{j_{\ell-1},r}(D)$ and then it samples $j_\ell$ from that distribution. We assume that it does this sampling by performing independent unbiased coin flips, and using them to simulate the distribution in question by a standard approach which uses with probability 1 finite number of flips. We call the sequence of the coin flips $r$. Since the length of $r$ is finite almost surely (one can sample from a countable-support distribution while almost surely using finitely many bits), the distribution of $r$ has countable support up to a set of measure zero; this will help us ensure that we have no issues with measurability. The value of $j_\ell$ then depends deterministically on $r$ and the values $M_{j_1,r}, \ldots, M_{j_{\ell-1},r}$. If $V$ releases some auxiliary information together with the values $M_{j_1,r}(D), \ldots, M_{j_{\ell-1},r}(D)$, it is only less likely to be private, and we may thus assume that $V$ also releases $r$.

Suppose we fix the randomness $r$ and let $U$ be the range of the mechanisms (if the ranges differ, we set $U$ to be their union). We define an equivalence relation on $U^k$ as follows. We say that $u, v \in U^k$ are equivalent if they result in the same sequence of $j_1, \ldots, j_k$ (for the fixed value of $r$). Let $s_1(r), s_2(r), \ldots$ be the equivalence classes under of this equivalence relation. The sequence $j_1, \ldots, j_k$ can then be seen as a function of $r$ and of $s_h(r)$ for $h \in \mathbb{N}$. The number of equivalence classes is countable for any $r$.

Let us have two neighboring databases $D_1 \sim D_2$. For any measurable $Y \subseteq U$ and $t \in \{0,1\}^*$, we can now bound the ratio

$$\frac{P(r = t \land (M_{j_1,r}(D_1), \ldots, M_{j_k,r}(D_1)) \in Y)}{P(r = t \land (M_{j_1',r}(D_2), \ldots, M_{j_k',r}(D_2)) \in Y)} = \frac{\sum_{h=1}^{\infty} P((M_{j_1,r}(D_1), \ldots, M_{j_k,r}(D_1)) \in Y \cap s_h(t)|P(r = t))}{\sum_{h=1}^{\infty} P((M_{j_1',r}(D_2), \ldots, M_{j_k',r}(D_2)) \in Y \cap s_h(t)|P(r = t))}$$

The values $j_\ell, j_\ell'$ in $P((M_{j_1,r}(D), \ldots, M_{j_k,r}(D)) \in Y \cap s_h|P(r = t))$ are deterministic as they are a function of $r$ and $s_h$, as noted above. As such, it holds $j_\ell = j_\ell'$. Therefore, by the assumption that $(M_{i_1,r}(D), \ldots, M_{i_k,r}(D))$ is $\epsilon$-differentially private for any fixed $i_1, \ldots, i_k$, it holds that

$$\frac{P((M_{j_1,r}(D_1), \ldots, M_{j_k,r}(D_1)) \in Y \cap s_h|P(r = t))}{P((M_{j_1',r}(D_2), \ldots, M_{j_k',r}(D_2)) \in Y \cap s_h|P(r = t))} \leq e^\epsilon$$

which allows us to bound $\epsilon$ as

$$\sum_{h=1}^{\infty} e^\epsilon P((M_{j_1',r}(D_2), \ldots, M_{j_k',r}(D_2)) \in Y \cap s_h|P(r = t)) = e^\epsilon$$

The other inequality holds by symmetry. This proves that $V$ is $\epsilon$-differentially private, as we wanted to prove.

We are now ready to prove the main theorem of this section. In what follows, we denote by $A_r(\cdot)$ the algorithm $A$ executed with randomness $r$. We formalize the multiple-query setting as having one algorithm which takes as part of its input a query. This differs somewhat from the presentation in the introduction which assumed we have a sequence of algorithms, which we chose there as it required less notation. Note also that while we are not assuming that the algorithm does
not know which phase it is (the value of \( i \)), we may without loss of generality assume this is passed as part of the query. Note that the condition on the queries \( x_1, \ldots, x_k \) below simply states that the queries can be chosen adaptively based on the released values, and that they do not have to be deterministic. Note also that the randomness \( r \) must not be released, as the privacy also relies on that randomness.

**Theorem 10.** Let us have an algorithm \( A(D, x) \) for a database \( D \) and a query \( x \in U \), where \( U \) is a countable universe. Assume there exists a function \( g(D, x) \) with its global sensitivity w.r.t. \( D \) being \( \leq \Delta_1 \) for any \( x \), and such that \( \text{Var}[A(D, x) - g(D, x)] \leq \Delta_2 \) for any \( D, x \).

Pick at random \( Y_i \sim \text{Laplace}(c(\Delta_1 + \Delta_2)/\epsilon) \) for \( c = 3 + 12 \log 2 \) and for \( \epsilon \leq c/6 \) and pick \( r \) independently uniformly on \( \{0, 1\}^\infty \). Then for queries \( x_1, \ldots, x_k \in U \) where the query \( x_i \) is drawn from a distribution that is a function of \( A_r(D, x_1) + Y_1, \ldots, A_r(D, x_{i-1}) + Y_{i-1} \), releasing \( A_r(D, x_1) + Y_1, \ldots, A_r(D, x_k) + Y_k \) is \( \epsilon \)-differentially private, with the privacy also being over the randomness of \( r \).

If \( k = 1 \), then \( c = 1 + 4 \log 2 \) and \( \epsilon \leq c/2 \) is sufficient.

**Proof.** Let us have two neighboring databases \( D_1, D_2 \). Let \( Y = (Y_1, \ldots, Y_k) \), and for \( x = (x_1, \ldots, x_k) \), let \( A'(D, x) = (A_r(D, x_1) + Y_1, \ldots, A_r(D, x_k) + Y_k) \) for \( r \) uniform on \( \{0, 1\}^\infty \). Similarly, let \( g(D, x) = (g(D, x_1), \ldots, g(D, x_k)) \). We prove that for any fixed (non-adaptive) queries \( x = (x_1, \ldots, x_k) \) it holds \( f_{A' \delta(D_1, x)}(y)/f_{A' \delta(D_2, x)}(y) \leq e^{\epsilon} \). If we prove this, the theorem follows: the inequality \( f_{A \delta(D_1, x)}(y)/f_{A \delta(D_2, x)}(y) \leq e^{\epsilon} \) holds by symmetry and these bounds together imply \( \epsilon \)-differential privacy. Lemma \( \text{[9]} \) then imply that \( A' \) is differentially private even for adaptive queries.

Let \( R_1 = A'(D_1, x) - g(D_1, x) \) and \( R_2 = A'(D_2, x) - g(D_1, x) \) (note the asymmetry in the definitions). We are assuming \( R_1 \) has subexponential diameter \( \leq \Delta_2 \). By Fact \( \text{[5]} \) it holds that \( R_2 \) has subexponential diameter \( \leq \Delta_2/\log(2) + \Delta_2 \). We now have from Lemma \( \text{[8]} \) the following bounds

\[
\frac{f_{g(D_1, x) + R_1 + Y}(y)}{f_{g(D_1, x) + Y}(y)} \leq e^{\epsilon \log(2)/c}
\]

\[
\frac{f_{g(D_2, x) + R_2 + Y}(y)}{f_{g(D_1, x) + Y}(y)} \geq e^{-3(1+\log 2)/c}
\]

which in turn allows us to bound

\[
f_{A'(D_1, x)}(y)/f_{A'(D_2, x)}(y) = \frac{f_{g(D_1, x) + R_1 + Y}(x)}{f_{g(D_1, x) + Y}(x)} \cdot \frac{f_{g(D_1) + Y}(x)}{f_{g(D_2) + R_2 + Y}(x)} \leq e^{(3+12 \log 2)/c} = e^{\epsilon}
\]

If \( k = 1 \), the same computation gives the desired bound for \( c = 1 + 4 \log 2 \), since Lemma \( \text{[8]} \) in that case gives gives

\[
\frac{f_{g(D_1, x) + R_1 + Y}(y)}{f_{g(D_1, x) + Y}(y)} \leq e^{3 \log(2)/c}
\]

\[
\frac{f_{g(D_2, x) + R_2 + Y}(y)}{f_{g(D_1, x) + Y}(y)} \geq e^{-(1+\log 2)/c}
\]

which again results in the bound \( f_{A'(D_1, x)}(y)/f_{A'(D_2, x)}(y) \leq e^{\epsilon} \).  

### 4 Algorithms with bounded mean error

In this section, we show a weaker version of Theorem \( \text{[10]} \) that only requires that the error has some number of bounded moments, instead of requiring that it is subexponential. We start by defining the distribution that we will use in our additive noise mechanism.
Definition 11. Zero-symmetric Pareto distribution with shape parameter $\alpha > 1$ and scale parameter $s > 0$, denoted $ZSPareto_\alpha(s)$, is defined by the PDF
\[
\frac{1}{2s}(\alpha - 1)(|x|/s + 1)^{-\alpha}
\]

It should be noted that the “scale parameter” $s$ indeed in some sense represents the scale of the distribution. Specifically,
\[
P[|ZSPareto_\alpha(s)| \leq t] = \int_{-t}^{t} \left( \frac{\alpha - 1}{s} \right) (\frac{|x|}{s} + 1)^{-\alpha} \, dx = 1 - \left(\frac{s + t}{s}\right)^{1-\alpha}
\]
which can be made arbitrarily close to 1 while setting $t = \Theta(s)$, meaning that an arbitrarily large fraction of the probability mass is within $O(s)$ of the origin.

Before we can prove the main theorem of this section, we need to prove a technical lemma.

Lemma 12. Let us have any $0 \leq \epsilon \leq 1$, $\alpha > 1$, and $x \geq 0$. It holds
\[
\int_0^1 \min\left((1 + x)\alpha, \left|1 - \frac{(1 - 2^{-1/\alpha})\epsilon}{(1 + x)(1 - u)^{1/\alpha}}\right|^{-\alpha}\right) du \leq 1 + \frac{2\alpha - 1}{\alpha - 1} \epsilon
\]

Proof. We start by proving that for $0 \leq u \leq 1 - \epsilon/(1 + x)^\alpha$ and $0 < \epsilon < 1$, $\alpha > 1$ and $x \geq 0$ holds that
\[
\left|1 - \frac{(1 - 2^{-1/\alpha})\epsilon}{(1 + x)(1 - u)^{1/\alpha}}\right|^{-\alpha} \leq 1 + \frac{\epsilon}{(1 - u)^{1/\alpha}}
\]
We express the inequality using $a$, defined by $u = 1 - a \frac{\epsilon}{(1 + x)^\alpha}$. The condition $u \leq 1 - \frac{\epsilon}{(1 + x)^\alpha}$ will then be equivalent to $a \geq 1$. It is thus sufficient to prove that for $a \geq 1$ it holds
\[
\left|1 - \frac{(1 - 2^{-1/\alpha})\epsilon}{a^{1/\alpha}}\right|^{-\alpha} \leq 1 + \frac{(1 + x)\epsilon}{a^{1/\alpha}}
\]
The right-hand side increases with $x$ while the left-hand side is independent of it, and we may thus set $x = 0$ (note that we are assuming $x \geq 0$). It holds that $0 \leq \epsilon^{1-1/\alpha}/a^{1/\alpha} \leq 1$ as $a \geq 1$ and $\epsilon < 1$. The inside in the absolute value is thus non-negative, and we may ignore the absolute value. It thus suffices to prove
\[
\left(1 - \frac{(1 - 2^{-1/\alpha})\epsilon}{a^{1/\alpha}}\right)^{-\alpha} \leq 1 + \frac{\epsilon^{1-1/\alpha}}{a^{1/\alpha}}
\]
As we said, it holds $0 \leq \epsilon^{1-1/\alpha}/a^{1/\alpha} \leq 1$. By further substituting $b = \epsilon^{1-1/\alpha}/a^{1/\alpha}$, proving our inequality reduces to proving
\[
(1 - (1 - 2^{-1/\alpha})b)^{-\alpha} \leq 1 + b
\]
under the condition $0 \leq b \leq 1$. The left-hand side of \(6\) is convex as a function of $b$ on $[0,1]$: the second derivative is
\[
\frac{(2^{1/\alpha} - 1)^2 \alpha(\alpha + 1)(1 - (1 - 2^{-1/\alpha})b)^{-\alpha}}{(-2^{1/\alpha} + 2^{1/\alpha}b - b)^2}
\]
which is non-negative as the denominator \((-2^{1/\alpha} + 2^{1/\alpha}b - b)^2\) is clearly non-negative, while the numerator is a product of \((2^{1/\alpha} - 1)^2\) which is non-negative, of \(\alpha(\alpha + 1)\) which is also non-negative, and of \((1 - (1 - 2^{-1/\alpha}) b)^{-\alpha}\) which is also non-negative as \(b \leq 1\) and thus \((1 - 2^{-1/\alpha}) b < 1\).

The right-hand side of (6) is an affine function of \(b\). It is thus sufficient to prove (6) for \(b = 0\) and \(b = 1\). The inequality clearly holds for \(b = 0\). For \(b = 1\), we have \((1 - (1 - 2^{-1/\alpha})b)^{-\alpha} = 2 = 1 + b\), thus proving inequality (6).

We now bound the integral. It holds

\[
\int_0^1 \min \left( (1 + x)^\alpha, 1 - \frac{(1 - 2^{-1/\alpha})e}{(1 + x)(1 - u)^{1/\alpha}} \right)^{-\alpha} du \leq \int_0^{1/\epsilon(1 + x)^\alpha} \left( 1 - \frac{(1 - 2^{-1/\alpha})e}{\alpha(1 + x)(1 - u)^{1/\alpha}} \right)^{-\alpha} du \\
+ \frac{\epsilon}{(1 + x)^\alpha} (1 + x)^\alpha \\
\leq \int_0^1 1 + \frac{\epsilon}{(1 - u)^{1/\alpha}} du + \epsilon \\
= 1 + \epsilon + \left[ \frac{\alpha u^{\frac{\alpha - 1}{\alpha}}}{\alpha - 1} \right]_{u=0}^{1} \epsilon = 1 + \frac{2\alpha - 1}{\alpha - 1} \epsilon
\]

where the second inequality uses inequality (5).

In what follows, we use the notation \(\|X\|_p\) for a random variable \(X\) to denote the \(L_p\) norm \(\sqrt[p]{\mathbb{E}[|X|^p]}\).

**Lemma 13.** Let \(X\) be a real random variable such that \(\|X\|_\alpha \leq \Delta\) and let \(Y \sim ZSPareto_\alpha(s)\) for \(s = (1 - 2^{-1/\alpha})^{-1}\Delta/\epsilon\) for \(0 \leq \epsilon \leq 1\) and \(\alpha > 1\). Consider \(y \in \mathbb{R}\). It holds \(e^{-(1 - 2^{-1/\alpha})\alpha \epsilon} \leq f_{X+Y}(y)/f_Y(y) \leq e^{\frac{2\alpha - 1}{\alpha - 1} \epsilon}\).

**Proof.** Since \(\|X\|_\alpha \leq \Delta\), it holds by the higher-order Chebyshev inequality\(^7\) that \(\mathbb{P}(|X| \geq z) \leq \Delta^\alpha / z^\alpha\). Therefore, \(F_{|X|}^{-1}(u) \leq \Delta / \sqrt[\alpha]{1 - u}\). We will use that \(|y - X| \leq |y| + |X|\) and later below also that \(|y - X| \geq \max(0, |y| - |X|)|. We start with the simpler case of proving a lower bound.

\[
\frac{f_{X+Y}(y)}{f_Y(y)} = \frac{\mathbb{E}[f_Y(y - X)]}{\mathbb{E}[|X|/s + 1]^{-\alpha}} \\
\geq \mathbb{E} \left[ \left( \frac{|y|/s + |X|/s + 1}{|y|/s + 1} \right)^{-\alpha} \right] \\
= \mathbb{E} \left[ \left( 1 + \frac{|X|/s}{|y|/s + 1} \right)^{-\alpha} \right] \\
\geq \mathbb{E}[1 + |X|/s]^{-\alpha} \\
\geq \mathbb{E}[1 - \alpha|X|/s] \\
= 1 - \alpha \mathbb{E}[|X|/s] \\
\geq 1 - \frac{\alpha(1 - 2^{-1/\alpha})\Delta}{\Delta} \epsilon \geq e^{-(1 - 2^{-1/\alpha})\alpha \epsilon}
\]

\(^7\)The higher-order Chebyshev inequality states that for \(X\) being a real random variable, it holds \(\mathbb{P}(|X - E[X]| \geq t) \leq e^{t^\alpha/\mathbb{E}[|X - E[X]|^\alpha]}\) for any \(\alpha \geq 0\).
Theorem 14. Let us have an algorithm \( A(D) \) such that there exists a function \( g(D) \) with global sensitivity \( \leq \Delta_1 \) w.r.t. \( D \) for which for any input \( D \), it holds \( \| A(D) - g(D) \|_1 \leq \Delta_1^2 \) for some \( \alpha > 1 \). Let \( Y \sim Z \phi_{c, \alpha} \) with global \( \alpha \)-sensitivity \( \leq \Delta_1 \) w.r.t. \( D \). We now argue the last inequality; that will conclude the proof. If we set \( \alpha \) to be an integer, then \( \alpha - 1 \) is a natural number, and we have proven the inequality second to last in Lemma 12.

Proof. We are now ready to prove the main theorem of this section.

Theorem 14. Let us have an algorithm \( A(D) \) such that there exists a function \( g(D) \) with global sensitivity \( \leq \Delta_1 \) w.r.t. \( D \) for which for any input \( D \), it holds \( \| A(D) - g(D) \|_1 \leq \Delta_1^2 \) for some \( \alpha > 1 \). Let \( Y \sim Z \phi_{c, \alpha} \) with global \( \alpha \)-sensitivity \( \leq \Delta_1 \) w.r.t. \( D \). We now argue the last inequality; that will conclude the proof. If we set \( \alpha \) to be an integer, then \( \alpha - 1 \) is a natural number, and we have proven the inequality second to last in Lemma 12.

We are now ready to prove the main theorem of this section.

Theorem 14. Let us have an algorithm \( A(D) \) such that there exists a function \( g(D) \) with global sensitivity \( \leq \Delta_1 \) w.r.t. \( D \) for which for any input \( D \), it holds \( \| A(D) - g(D) \|_1 \leq \Delta_1^2 \) for some \( \alpha > 1 \). Let \( Y \sim Z \phi_{c, \alpha} \) with global \( \alpha \)-sensitivity \( \leq \Delta_1 \) w.r.t. \( D \). We now argue the last inequality; that will conclude the proof. If we set \( \alpha \) to be an integer, then \( \alpha - 1 \) is a natural number, and we have proven the inequality second to last in Lemma 12.

We are now ready to prove the main theorem of this section.

Theorem 14. Let us have an algorithm \( A(D) \) such that there exists a function \( g(D) \) with global sensitivity \( \leq \Delta_1 \) w.r.t. \( D \) for which for any input \( D \), it holds \( \| A(D) - g(D) \|_1 \leq \Delta_1^2 \) for some \( \alpha > 1 \). Let \( Y \sim Z \phi_{c, \alpha} \) with global \( \alpha \)-sensitivity \( \leq \Delta_1 \) w.r.t. \( D \). We now argue the last inequality; that will conclude the proof. If we set \( \alpha \) to be an integer, then \( \alpha - 1 \) is a natural number, and we have proven the inequality second to last in Lemma 12.

We are now ready to prove the main theorem of this section.

Theorem 14. Let us have an algorithm \( A(D) \) such that there exists a function \( g(D) \) with global sensitivity \( \leq \Delta_1 \) w.r.t. \( D \) for which for any input \( D \), it holds \( \| A(D) - g(D) \|_1 \leq \Delta_1^2 \) for some \( \alpha > 1 \). Let \( Y \sim Z \phi_{c, \alpha} \) with global \( \alpha \)-sensitivity \( \leq \Delta_1 \) w.r.t. \( D \). We now argue the last inequality; that will conclude the proof. If we set \( \alpha \) to be an integer, then \( \alpha - 1 \) is a natural number, and we have proven the inequality second to last in Lemma 12.
5 Implications of our results

In this section, we give several implications of Theorem 10 and Theorem 14. This list is by no means meant to be exhaustive. We start with the more straightforward applications and focus on the more involved ones later.

Recall that, as we discussed, the goal in the sublinear setting is not to simply add small amount of noise, but rather to achieve a given level of error as efficiently as possible while guaranteeing differential privacy. This is so because in this setting, the amount of error coming from the algorithm not being exact tends to be much greater than the amount of noise needed to achieve privacy when not subject to having limited resources.

5.1 The general approach

In all applications, we take a known algorithm for a given problem, and use either Theorem 10 or Theorem 14 to argue that adding noise to the algorithm’s answer ensures privacy.

Assume the original algorithm had complexity $T(n, \rho)$ and assume for example that the error is of magnitude $pm$, namely that for error $R$, it holds $E[R^2]^{1/2} \leq O(pm)$ (this can be generalized to $\leq O(pm(f(x)))$ for $x$ being the input and $f$ being any function). We run the algorithm with parameter $\rho' = \rho \epsilon$ and add noise of magnitude $O(pm)$ (more generally $O(pm(f(x)))$. By Theorem 14 with $\alpha = 2$, as long as the approximated function’s sensitivity is $\Delta \leq \epsilon \rho n$, this is $\epsilon$-differentially private. At the same time, the error is $\leq O(pm)$ with arbitrarily high constant probability. The time complexity is $T(n, \rho)$.

If we want to achieve a failure probability of $\beta$, we run this algorithm $\Theta(\log \beta^{-1})$ times and take the median. By a standard probability amplification argument, the success probability will be as desired. To achieve $\epsilon$-differential privacy by composition, we have to divide the privacy budget between the runs, resulting in complexity $O(T(n, \epsilon \rho / \log \beta^{-1}) \log \beta^{-1})$. This can be summarized (and generalized with the general function $f(x)$) as follows:

Lemma 15. Suppose there is an algorithm approximating a function $g$ with global sensitivity $\Delta$ such that $E[(A(x) - g(x))^2]^{1/2} \leq \rho f(x)$ for some function $f$ with time/space/query complexity $T(n, \rho)$. Then for $\epsilon \leq O(1)$ there exists an $\epsilon$-differentially private algorithm $A'$ such that when $\epsilon \rho \geq \Omega(\Delta / f(x))$, it holds $P[|A'(x) - g(x)| > \rho f(x)] \leq \beta$ and that has complexity $O(T(n, \epsilon \rho / \log \beta^{-1}) \log \beta^{-1})$.

A more efficient approach for decreasing failure probability exists in the case of subexponential error. Assume the same setting as above, except that the error’s subexponential diameter is $\rho n$ instead having only moment bounds (like above, we can generalize to $\rho f(x)$ instead of $\rho n$). We run the algorithm with parameter $\rho' = \epsilon \rho / \log \beta^{-1}$ and add noise of magnitude $\Theta(\rho n / \log \beta^{-1})$. This algorithm is $\epsilon$-differentially private by Theorem 10 as long as $\Delta \leq \epsilon \rho n$. The noise has subexponential diameter $O(\rho' n)$ and by Lemma 6, the total error will up to a constant have the same subexponential diameter. By the definition of subexponential diameter, the probability that the error is $\geq \Theta(\rho n)$ is $\leq \beta$ as desired. This results in complexity $O(T(n, \epsilon \rho / \log \beta^{-1}))$ saving us one log $\beta^{-1}$ factor. We can summarize this as

Lemma 16. Suppose there is an algorithm approximating a function $g$ with global sensitivity $\Delta$ such that $\sigma_{\epsilon}(A(x) - g(x)) \leq \rho f(x)$ for some function $f$ with time/space/query complexity $T(n, \rho)$. Then

---

8 This upper bound on the sensitivity ensures that $\Delta_2$ in Theorem 14 dominates.
9 The error coming from the algorithm will not be too large by Chebyshev and the error from the noise will not be too large by our bound 8. By the union bound, neither of the two sources of error will be too large with arbitrarily high constant probability.
for $\epsilon \leq O(1)$, there exists an $\epsilon$-differentially private algorithm $A'$ such that when $\epsilon \rho \geq \Omega(\Delta/f(x))$, it holds $P[|A'(x) - g(x)| > \rho f(x)] \leq \beta$ and that has complexity $O(T(n, \epsilon \rho/ \log \beta^{-1}))$.

We are now ready to give private algorithms for specific problems.

5.2 Frequency moment $F_2$

In their seminal paper, Alon, Matias, and Szegedy [2] show a sketch that allows one to estimate the $F_2$ frequency moment, defined as $F_2(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i^2$. In the streaming setting, the vector $x_1, \ldots, x_n$ is given through a stream of updates $y_1, \ldots, y_k$ of the form $y_j = (\ell_j, D_j)$ where $D$ can be negative and we define $x_i = \sum_{j=1}^{k} I[\ell_j = i]D_j$. Two inputs are then adjacent if they differ in one value $y_j$ for some $j$. The algorithm from [2] uses space $O(\frac{1}{\rho^2})$ and has mean squared error of $\leq \rho^2 F_2^2 \leq \rho^2 n^4$. The sensitivity of the $F_2$ moment is $n$. This implies the following

**Corollary 17.** For $\epsilon \leq O(1)$ and $\rho \leq 1/(en)$, there is an $\epsilon$-differentially private algorithm that returns an additive $\pm \rho n^2$ approximation of the frequency moment $F_2$ with probability $1 - \beta$, and has space complexity $O(\frac{\log^3 \beta^{-1}}{\rho^2 \epsilon^2})$.

No streaming algorithm for private $F_2$ moments was previously known. Shortly after releasing this paper, an approach with the incomparable complexity of $O(\log(n) \log^2 \beta^{-1}/\rho^2)$ was shown (for $\epsilon$ being not too small) that also has multiplicative approximation guarantees and works (with some loss in the complexity) in the continual release setting [16]. The setting of $F_p$ for $p \in [0, 1]$ has been considered in [35].

5.3 Connected components

Berenbrink, Kratennhoff, and Mallmann-Trenn [5] have shown an algorithm that returns an estimate $\hat{c}$ of the number of connected components $c$ of a simple graph in time $O(\frac{1}{\rho^2} \log \frac{1}{\rho})$ and has mean squared error $\mathbb{E}[(\hat{c} - c)^2] \leq \rho^2 n^2$. At the same time, the number of connected components has global sensitivity $1$ with respect to edge additions/deletions. This gives us the following

**Corollary 18.** For $\epsilon \leq O(1)$ and $\rho \leq 1/(en)$, there is an $\epsilon$-edge-differentially private algorithm that returns an additive $\pm \rho n$ approximation of the number connected components with probability $1 - \beta$, and has complexity $O(\frac{\log^2 \beta^{-1}}{\rho^2 \epsilon^2} \log \frac{\log \beta^{-1}}{\rho \epsilon})$.

No private sublinear-complexity algorithm for estimating the number of connected components was previously known.

5.4 Maximum matching

Yoshida et al. [36] show an algorithm that can approximate the size of the maximum matching to within multiplicative $1 + \rho$ in time $d^{O(1/\rho^2)}(1/\rho)^{O(1/\rho)}$ for $d$ being the maximum degree of the input graph. It works by implementing an oracle for a matching of size within factor $1 + \rho/2$ of the maximum matching; for a specified vertex, this oracle answers whether the vertex is matched in the oracle’s matching. The algorithm then samples $\Theta(1/\rho^2)$ vertices and checks the fraction that is matched in the oracle’s matching. The error coming from the oracle is $\leq \rho n^2/2$ in the worst case and thus has subexponential diameter $O(\rho n)$. The error coming from the sampling has subexponential diameter $O(\rho n)$ by the Hoeffding inequality. By Lemma [6] the subexponential diameter of the error is thus $O(\rho n)$. At the same time, the global sensitivity of the maximum matching size is $\leq 1$ with respect to the removal of one vertex. This gives us the following
Corollary 19. For $\epsilon \leq O(1)$ and $\rho \leq 1/(en)$, there is an $\epsilon$-node-differentially-private algorithm that returns an additive $\pm \rho n$ approximation of the maximum matching size with probability $1 - \beta$ in time $d^{O((\log(\beta^{-1}))/((\rho^2)^2))}/(p\epsilon)O((\log(\beta^{-1}))/((\rho^2)))$.

This solves the open problem posed in [6] where the authors show a hybrid $(2, \rho n)$ approximation, while we give a purely additive $\pm \rho n$ approximation.

5.5 Rank queries

Karnin et al. [21] develop a sketch of size $O(\frac{1}{\rho})$ that allows one to answer rank queries with error with subexponential diameter $\rho n$. We show how to use their sketch to answer range queries over an ordered universe. For small number of queries, this improves upon the work of [20] which has a logarithmic dependency on the universe size. This gives us the following corollary.

Corollary 20. There is a sketch that allows $\epsilon$-differentially private algorithm that returns k rank queries (potentially adaptive) for $\epsilon \leq O(1)$ with an additive $\pm \rho n$ error with probability $1 - \beta$, and has complexity $O(\frac{k \log^2(k/\beta)}{\rho \epsilon})$.

Proof. We use the KLL sketch with error parameter $\rho' = \rho \epsilon/(k \log(k/\beta))$. This means that the error has a subexponential diameter of $\leq \rho n/(k \log(k/\beta))$. Therefore, by Theorem 10, it holds that using for each query a Laplace mechanism with error magnitude $\Theta(\rho n/\log(k/\beta))$ will result in $\epsilon$-differential privacy.

By Lemma 6 the overall subexponential diameter the error of each answer is $O(\rho n/\log(k/\beta))$ and there the probability of error $O(\rho n)$ is $1 - \beta/k$. By the union bound, the overall success probability is $\geq 1 - \beta$.

This is in comparison to the approach of Kaplan and Stemmer [20] which results in space complexity $O(\frac{\log |U| \log(k/\beta)}{\rho \epsilon})$, improving by a factor of $\log |U|$ for constant $k, \beta$, where $U$ is the universe.

5.6 Relative approximation sublinear-time algorithms

A common way of designing sublinear time algorithms with relative error is to design an algorithm that requires “advice” in the form of a constant-factor approximation of the answer and then getting rid of the need for this advice [15]. While this may seem impossible, given that the algorithm needs to roughly know the correct answer in order to give a correct estimate, there is a way that makes this possible under some assumptions. We now show that this advice-removal technique can be modified to also handle differential privacy under mild assumptions.

First, we need a simple lemma.

Lemma 21. Let $X_1, \cdots, X_{2k-1}$ be i.i.d. non-negative random variables. Then

$$\mathbb{E}[\text{median}(X_1, \cdots, X_{2k-1})] \leq 2 \left(\frac{2k - 1}{k}\right) \mathbb{E}[X_1]$$

Proof. $\mathbb{E}[\text{median}(X_1, \cdots, X_{2k-1})] = \int_0^\infty \mathbb{P}[\text{median}(X_1, \cdots, X_{2k-1}) \geq t] dt$

$$\leq \int_0^\infty \left(\frac{2k - 1}{k}\right) \mathbb{P}[X_1 \geq t]^k dt$$

$$\leq \left(\frac{2k - 1}{k}\right) \int_0^\infty \min(1, \mathbb{E}[X_1]^k/t^k) dt$$
Then for $\epsilon$ that

$$\text{Theorem 22. Let us have an algorithm where the first inequality is by union bound, over all subsets of size } k \text{ and the second inequality holds by the Markov inequality.}$$

We are now ready to prove the theorem. The value $y$ in the following is the advice and intuitively speaking, one should think that we want $y$ to be $\approx g(x)$.

**Theorem 22.** Let us have an algorithm $A(x, y, \rho)$, a function $g(x)$ with global sensitivity $\Delta$ such that $\sup_x g(x) \leq M$ and $\inf_x g(x) \geq m$. Assume that

1) $\mathbb{E}[A(x, y, \rho)] \leq g(x)$,

2) $\mathbb{E}[|A(x, y, \rho) - g(x)|] \leq \rho y$,

3) has complexity $O(T(x, y, \rho))$ such that $O(T(x, y', \rho) \cdot (y'/y)^c_1) \leq T(x, y, \rho) \leq O(T(x, y', \rho) \cdot (y'/y)^c_2)$ for any $x, y', y, \rho$ and some constants $c_1, c_2 > 0$.

Then for $\epsilon \leq O(1 / \log(M/n))$ and $\rho \leq \Delta/\epsilon(n)$, there is an $\epsilon$-differentially private algorithm $A'(x)$ such that $\mathbb{P}[|A'(x) - g(x)| \geq \Theta(\rho g(x))] \leq 1/3$ with complexity $O(T(x, g(x), \epsilon/\log(M/m)) + T(x, g(x), \epsilon\rho))$.

**Proof.** Let us denote with $B(x, y, \rho)$ independently executing $A(x, y, \rho)$ five times and taking the median. By the moment amplification lemma (Lemma 3), we have that

$$\mathbb{E}[|B(x, y, \rho) - g(x)|^3] \leq 10^{3} y^{3}$$

Therefore, by Theorem 14 $B(x, y, \rho) + ZCPareto_{3}(55\rho y/\epsilon)$ is $\epsilon$-differentially private.

We now describe the algorithm $A'$. We set $\tilde{y} = M$, and in each iteration, we will decrease $\tilde{y}$ by a factor of 2. In each iteration, we compute $B(x, \tilde{y}, 330^{-3/2} \log_{2}(M/m)) + ZCPareto_{3}(\frac{\tilde{y}}{3^{2}})$. When the returned value is $\geq \tilde{y}$, we compute and return $B(x, \tilde{y}/160, \rho\epsilon) + ZCPareto_{3}(\rho\tilde{y})$.

The algorithm is $\epsilon$-differentially private by composition, as we executed at most $\log_{2}(M/m)$ times an algorithm that was $\epsilon/(2 \log_{2}(M/m))$ differentially private, and once an algorithm that was $\epsilon/2$-differentially private.

We now argue correctness. It holds $\mathbb{E}[ZCPareto_{3}(55\rho y/\epsilon)] = 0$ for any $\epsilon$, and by Lemma 21 we have $\mathbb{E}[B(x, y, \rho)] \leq 20g(x)$. Therefore,

$$\mathbb{E}[B(x, y, \rho) + ZCPareto_{3}(55\rho y/\epsilon)] \leq 20g(x)$$

for any $\epsilon$. By the Markov inequality, the probability that $B(x, \tilde{y}, 330^{-3/2} \log_{2}(M/m)) + ZCPareto_{3}(\frac{\tilde{y}}{3^{2}}) \geq z$ is then at most $20g(x)/z$. The probability that we stop with $y \geq 160g(x)$ is thus at most

$$\sum_{i=1}^{\lceil \log_{2}(M/m) \rceil} \mathbb{I}[M/2^i \geq 160g(x)] \mathbb{P}[B(x, y, \rho) + ZCPareto_{3}(55\log_{2}(M/m)/\epsilon) \geq M/2^i]$$

18
\[
\leq \sum_{j=0}^{\infty} \mathbb{P}[B(x, y, \rho) + ZCPareto_3(55 \log_2(M/m)/\epsilon) \geq 160 \cdot 2^j g(x)] \\
\leq \sum_{j=0}^{\infty} 2^{-j/8} = 1/4
\]

Therefore, with probability at least 3/4, we have in the last iteration that \( \tilde{y}/160 \leq g(x) \). We call this event \( E \). On \( E \), we have

\[
\mathbb{E}[|B(x, \tilde{y}/160, \epsilon \rho)| + ZCPareto_3(\epsilon \rho y) - g(x)|] \leq \mathbb{E}[|B(x, \tilde{y}/160, \epsilon \rho)| - g(x)|] + \mathbb{E}[|ZCPareto_3(\epsilon \rho y)|]
\leq O(\epsilon \rho g(x)) + O(\rho g(x)) = O(\rho g(x))
\]

where the first term comes from the assumption 2) and the second is because \( \mathbb{E}[|ZCPareto_3(\epsilon \rho y)|] = a \)

On \( E \), we thus have by the Markov inequality that the error is \( O(\rho g(x)) \) with probability \( 1 - 1/12 \) (or any probability arbitrarily close to 1). Adding the probability of \( \neg E \), we have that the error is \( O(\rho g(x)) \) with probability \( \geq 2/3 \), as we wanted to prove.

It remains to argue time complexity. Suppose \( \tilde{y} \leq g(x) \). Then

\[
\mathbb{E}[|B(x, \tilde{y}, \epsilon \rho y, 330 \cdot 3^{3c} \log_2(M/m)| + ZCPareto_3(\tilde{y}/(3 \cdot 3^{3c})) - g(x)|]
\leq \mathbb{E}[|B(x, \tilde{y}, \epsilon \rho y, 330 \cdot 3^{3c} \log_2(M/m)| - g(x)|] + \mathbb{E}[|ZCPareto_3(\tilde{y}/(3 \cdot 3^{3c}))|]
\leq \frac{g(x)}{330 \cdot 3^{3c}} + \frac{g(x)}{3 \cdot 3^{3c}} < \frac{g(x)}{2 \cdot 3^{3c}}
\]

and by the Markov inequality, it thus holds

\[
\mathbb{P}[B(x, \tilde{y}, \epsilon \rho y, 330 \cdot 3^{3c} \log_2(M/m)| + ZCPareto_3(\tilde{y}/(3 \cdot 3^{3c})) \geq g(x)/2] \geq 1 - 1/3^{3c}
\]

This means that if \( \tilde{y} \leq g(x)/2 \), we stop in each iteration with probability \( \geq 1 - 1/3^{3c} \). If \( \ell \) denotes the number of iterations, then \( \mathbb{P}[\ell \geq \lceil \log_2(2M/g(x)) \rceil + i] \leq 3^{-c_i} \). The expected complexity of the executions of \( B(x, \tilde{y}, \epsilon \rho y, 330 \cdot 3^{3c} \log_2(M/m)) \) is thus

\[
\sum_{i=0}^{\log(M/m)} T(x, M/2^i, \epsilon \rho y, 330 \cdot 3^{3c} \log_2(M/m)) \mathbb{P}[\ell \geq i] \leq \sum_{i=0}^{\lceil \log_2(M/g(x)) \rceil} T(x, M/2^i, \epsilon \rho y, 330 \cdot 3^{3c} \log_2(M/m))
+ \sum_{i=\lceil \log_2(2M/g(x)) \rceil}^{\infty} (3c)^{-i} T(x, M/2^i, \epsilon \rho y, 330 \cdot 3^{3c} \log_2(M/m))
\leq \sum_{j=0}^{\infty} T(x, g(x)/2^j/2, \epsilon \rho y, 330 \cdot 3^{3c} \log_2(M/m))
+ \sum_{j=0}^{\infty} T(x, g(x)/2^j, \epsilon \rho y, 330 \cdot 3^{3c} \log_2(M/m)) \mathbb{P}[\ell \geq i]
\leq \sum_{j=0}^{\infty} T(x, g(x), \epsilon \rho y, 330 \cdot 3^{3c} \log_2(M/m)) 2^{-c_i(j-1)}
\]

\[\text{This can be easily checked: } \mathbb{E}[|ZCPareto_3(a)|] = \int_0^\infty \mathbb{P}[|ZCPareto_3(a)| \geq t] dt = \left[-(a+t)(\frac{a+t+1}{a})^{2t+a}\right]_{t=0}^{\infty} = a.\]
\[ + \sum_{j=0}^{\infty} T(x, g(x), \frac{\epsilon}{330 \cdot 3^{2j} \log_2(M/m)}) 2^{c_1 \cdot 3 - c_2 j} \]

By the exact same argument, the executions of \( B(x, \tilde{y}/160, \epsilon \rho) \) contribute \( O(T(x, g(x), \epsilon \rho)) \), as the computation above is correct even if \( T(x, M/2^i, \frac{\epsilon}{330 \cdot 3^{2j} \log_2(M/m)}) \) is replaced by \( T(x, M/(160 \cdot 2^i), \rho \epsilon) \).

This concludes the proof.

This implies a differentially private algorithm for estimating the average degree of a graph. As the algorithm \( A \) of Theorem \[22\] we use the algorithm of Seshadhri \[30\]. This improves upon the work of Blocki et al. \[6\] who give an algorithm with complexity \( \tilde{O}(\epsilon, \rho, \sqrt{n}) \) under the assumption \( m \geq \Omega(n) \).

**Corollary 23.** For \( \epsilon \leq O(1) \) and \( \rho \leq 1/(\epsilon n) \), there exists an \( \epsilon \)-edge-differentially private algorithm that returns a \( 1 + \rho \)-approximation of the average degree of a graph with probability \( 1 - \beta \) and has complexity \( \tilde{O}(\frac{n \log^3 \beta^{-1}}{\epsilon^2 \rho^2 \sqrt{m}}) \).

**Proof.** As the authors prove, their estimator is unbiased. The variance of their algorithm is analyzed in \[3, Section 3.2\], proving that the variance is \( O(\rho^2 m^{3/2} \sqrt{y}) \) and complexity \( O(n/(\rho^2 \sqrt{y})) \) for \( y \) being an “advice parameter” like in Theorem \[22\]. Using Theorem \[22\] gives us a private with constant success probability and complexity \( \tilde{O}(\frac{n}{\epsilon^2 \rho^2 \sqrt{m}}) \). By standard probability amplification (dividing the privacy budget between the executions), we get the desired claim. \( \square \)

### 6 Open problems and conjectures

**Improve constants for the noise magnitude.** It seems likely that our analysis is not tight in terms of the constants. Is it possible to improve them? What are the best possible constants? Can the constants be improved for subgaussian error? Is it possible to get improved constants for some specific problems by using a definition of a subexponential diameter which differs by a constant factor\(^{11}\)? Is it possible to improve the constants in the subexponential case, if we assume that the estimator is unbiased (that is, that \( \mathbb{E}[A(D, x)] = g(D, x) \))? Can the constants be improved in the case of polynomial tails of the error? Are there any other distributions than the zero-symmetric power-law mechanism, which would achieve differential privacy under the same assumptions, but would have smaller variance/mean absolute deviation?

**Lower bounds for multiple queries with bounded error moments.** Suppose we want to publish the values of \( k \) dependent random variables, like in the subexponential case, but the error only has a finite number of finite moments. Does there exist a distribution of noise such that adding this noise to the output of the algorithm results in differential privacy? We conjecture that this is not possible and that it is inherent that the zero-symmetric Pareto mechanism only works for a number of variables that depends on the number of finite moments of the error. If this is the case, what is the number of moments, that we need (as a function of \( k \)) for a multivariate variant of Theorem \[14\] to hold? If this is some slow-growing function of \( k \), it may be feasible to use the median trick to bound higher moments like in \[22\], thus allowing a weaker multivariate version of Theorem \[14\] when only assuming an upper bound on the mean absolute deviation, but with superlinear dependency on \( k \).

\(^{11}\) As we mentioned in Section \[2\] there are several definitions of \( \sigma_{se} \), that differ by constant factors. See \[34\] for the definitions.
Normal noise in the case of subgaussian error. Suppose the algorithm’s error distribution is subgaussian (and not just subexponential). Is it possible to use, in that case, the Gaussian mechanism, and get the error scale with $\sqrt{k}$ instead of with $k$ for $k$ being the number of queries?

What if the amount of error depends on the input? There are some cases in which our approach does not apply. One possible reason is that the amount of error is large in the worst case, but it may be small in a certain instance. Perhaps one could use something akin to smoothed sensitivity? One notable case where this problem arises is the relative estimation of the frequency $F_p$ moments in the streaming setting. In that problem, the amount of error depends on the actual $F_p$ norm, and as such an instance-independent upper bound on the amount of error may not capture well the actual amount of error.

Acknowledgements

I would like to thank Rasmus Pagh and anonymous reviewers for helping to improve this paper.

References

[1] Daniel Alabi, Omri Ben-Eliezer, and Anamay Chaturvedi. Bounded space differentially private quantiles. arXiv preprint arXiv:2201.03380, 2022.

[2] Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. In Proceedings of the twenty-eighth annual ACM symposium on Theory of computing, pages 20–29, 1996.

[3] Sepehr Assadi. Sublinear time algorithms: Connected components, average degree. Advanced Algorithms II – Sublinear Algorithms, (Lecture 2), 2020.

[4] Raef Bassily, Kobbi Nissim, Uri Stemmer, and Abhradeep Guha Thakurta. Practical locally private heavy hitters. Advances in Neural Information Processing Systems, 30, 2017.

[5] Petra Berenbrink, Bruce Krayenhoff, and Frederik Mallmann-Trenn. Estimating the number of connected components in sublinear time. Information Processing Letters, 114(11):639–642, 2014.

[6] J Blocki, E Grigorescu, and T Mukherjee. Privately estimating graph parameters in sublinear time. In 49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)., 2022.

[7] Jeremiah Blocki, Avrim Blum, Anupam Datta, and Or Sheffet. The johnson-lindenstrauss transform itself preserves differential privacy. In 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science, pages 410–419. IEEE, 2012.

[8] Jeremiah Blocki, Elena Grigorescu, Tamalika Mukherjee, and Samson Zhou. How to make your approximation algorithm private: A black-box differentially-private transformation for tunable approximation algorithms of functions with low sensitivity. arXiv preprint arXiv:2210.03831, 2022.

[9] Jonas Boehler and Florian Kerschbaum. Secure sublinear time differentially private median computation, February 1 2022. US Patent 11,238,167.
[10] Seung Geol Choi, Dana Dachman-Soled, Mukul Kulkarni, and Arkady Yerukhimovich. Differentially-private multi-party sketching for large-scale statistics. Cryptology ePrint Archive, 2020.

[11] Charlie Dickens, Justin Thaler, and Daniel Ting. (nearly) all cardinality estimators are differentially private. arXiv preprint arXiv:2203.15400, 2022.

[12] Marianne Durand and Philippe Flajolet. Loglog counting of large cardinalities. In Algorithms-ESA 2003: 11th Annual European Symposium, Budapest, Hungary, September 16-19, 2003. Proceedings 11, pages 605–617. Springer, 2003.

[13] Cynthia Dwork. Differential Privacy. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, Automata, Languages and Programming, pages 1–12, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. ISBN 978-3-540-35908-1.

[14] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In Theory of cryptography conference, pages 265–284. Springer, 2006.

[15] Talya Eden, Dana Ron, and C Seshadhri. On approximating the number of k-cliques in sublinear time. In Proceedings of the 50th annual ACM SIGACT symposium on theory of computing, pages 722–734, 2018.

[16] Alessandro Epasto, Jieming Mao, Andres Munoz Medina, Vahab Mirrokni, Sergei Vassilvitskii, and Peilin Zhong. Differentially private continual releases of streaming frequency moment estimations. arXiv preprint arXiv:2301.05605, 2023.

[17] geetha290krm (https://math.stackexchange.com/users/1064504/geetha290krm). Does \( f_{X+Y}(z) = e[f_y(z-x)] \) hold? Mathematics Stack Exchange, 2022. URL:https://math.stackexchange.com/q/4544852 (version: 2022-10-04).

[18] Badih Ghazi, Noah Golowich, Ravi Kumar, Rasmus Pagh, and Ameya Velingker. On the power of multiple anonymous messages: Frequency estimation and selection in the shuffle model of differential privacy. In Annual International Conference on the Theory and Applications of Cryptographic Techniques, pages 463–488. Springer, 2021.

[19] Ziyue Huang, Yuan Qiu, Ke Yi, and Graham Cormode. Frequency estimation under multiparty differential privacy: One-shot and streaming. arXiv preprint arXiv:2104.01808, 2021.

[20] Haim Kaplan and Uri Stemmer. A note on sanitizing streams with differential privacy. arXiv preprint arXiv:2111.13762, 2021.

[21] Zohar Karnin, Kevin Lang, and Edo Liberty. Optimal quantile approximation in streams. In 2016 ieee 57th annual symposium on foundations of computer science (focs), pages 71–78. IEEE, 2016.

[22] Kasper Green Larsen, Rasmus Pagh, and Jakub Têtek. Countsketches, feature hashing and the median of three. In International Conference on Machine Learning, pages 6011–6020. PMLR, 2021.

[23] Christian Janos Lebeda and Jakub Têtek. Better differentially private approximate histograms and heavy hitters using the misra-gries sketch. arXiv preprint arXiv:2301.02457, 2023.
[24] Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press, 2015.

[25] Luca Melis, George Danezis, and Emiliano De Cristofaro. Efficient private statistics with succinct sketches. *arXiv preprint arXiv:1508.06110*, 2015.

[26] Darakhshan Mir, Shan Muthukrishnan, Aleksandar Nikolov, and Rebecca N Wright. Pan-private algorithms via statistics on sketches. In *Proceedings of the thirtieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 37–48, 2011.

[27] Jayadev Misra and David Gries. Finding repeated elements. *Science of computer programming*, 2(2):143–152, 1982.

[28] Rasmus Pagh and Mikkel Thorup. Improved utility analysis of private countsketch. *arXiv preprint arXiv:2205.08397*, 2022.

[29] Ryan M Rogers, Aaron Roth, Jonathan Ullman, and Salil Vadhan. Privacy odometers and filters: Pay-as-you-go composition. *Advances in Neural Information Processing Systems*, 29, 2016.

[30] C Seshadhri. A simpler sublinear algorithm for approximating the triangle count. *arXiv preprint arXiv:1505.01927*, 2015.

[31] Harry Sivasubramaniam, Haoran Li, and Xi He. Differentially private sublinear average degree approximation, 2020.

[32] Adam Smith, Shuang Song, and Abhradeep Guha Thakurta. The flajolet-martin sketch itself preserves differential privacy: Private counting with minimal space. *Advances in Neural Information Processing Systems*, 33:19561–19572, 2020.

[33] Nina Mesing Stausholm. Improved differentially private euclidean distance approximation. In *Proceedings of the 40th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, pages 42–56, 2021.

[34] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[35] Lun Wang, Iosif Pinelis, and Dawn Song. Differentially private fractional frequency moments estimation with polylogarithmic space. *arXiv preprint arXiv:2105.12363*, 2021.

[36] Yuichi Yoshida, Masaki Yamamoto, and Hiro Ito. An improved constant-time approximation algorithm for maximum matching. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 225–234, 2009.

[37] Fuheng Zhao, Dan Qiao, Rachel Redberg, Divyakant Agrawal, Amr El Abbadi, and Yu-Xiang Wang. Differentially private linear sketches: Efficient implementations and applications. *arXiv preprint arXiv:2205.09873*, 2022.
A Deferred proofs

Proof of Lemma 6

Proof. By the Markov inequality, we have for any $\lambda > 0$ that

$$P[\left| \sum_{i=1}^{k} X_i \right| \geq t] = P[\exp(\lambda \left| \sum_{i=1}^{k} X_i \right|) \geq e^{\lambda t}] \leq E[\exp(\lambda \left| \sum_{i=1}^{k} X_i \right|)] e^{-\lambda t} \quad (9)$$

If we set $\lambda = (3 \sum_{i=1}^{k} \sigma_{se}[X_i])^{-1}$, then we get

$$E \left[ \exp \left( \lambda \left| \sum_{i=1}^{k} X_i \right| \right) \right] = E \left[ \exp \left( \frac{\left| \sum_{i=1}^{k} X_i \right|}{3 \sum_{i=1}^{k} \sigma_{se}[X_i]} \right) \right]$$

$$\leq E \left[ \exp \left( \frac{\sum_{i=1}^{k} |X_i|}{3 \sum_{i=1}^{k} \sigma_{se}[X_i]} \right) \right]$$

$$= E \left[ \exp \left( \sum_{i=1}^{k} \frac{\sigma_{se}[X_i]}{3 \sum_{i=1}^{k} \sigma_{se}[X_i]} \frac{|X_i|}{3 \sigma_{se}[X_i]} \right) \right]$$

$$\leq \sum_{i=1}^{k} \frac{\sigma_{se}[X_i]}{3 \sum_{i=1}^{k} \sigma_{se}[X_i]} E \left[ \exp \left( \frac{|X_i|}{3 \sigma_{se}[X_i]} \right) \right] \leq (\ast)$$

where the last line is by the Jensen inequality. We have from Lemma 7 that $E[\exp(\frac{|X_i|}{3 \sigma_{se}[X_i]})] \leq 2^{1/3} \frac{1}{1/3} < 2$, where we have used that $\sigma_{se}[\frac{|X_i|}{3 \sigma_{se}[X_i]]} = 1/3$, giving us

$$(\ast) \leq \sum_{i=1}^{k} \frac{\sigma_{se}[X_i]}{\sum_{i=1}^{k} \sigma_{se}[X_i]} \cdot 2 = 2.$$  

This allows us to use (9) to bound

$$P[\left| \sum_{i=1}^{k} X_i \right| \geq t] \leq 2 e^{-t/(3 \sum_{i=1}^{k} \sigma_{se}[X_i])}$$

which is equivalent to $\sigma_{se}[\sum_{i=1}^{t} X_i] \leq 3 \sum_{i=1}^{k} \sigma_{se}[X_i]$ by the definition of $\sigma_{se}$. \qed