Proceedings of Machine Learning Research vol 178:1–37, 2022 35th Annual Conference on Learning Theory

A bounded-noise mechanism for differential privacy

Yuval Dagan DAGAN@CSAIL.MIT.EDU
MIT, EECS

Gil Kur GILKUR@MIT.EDU
MIT, EECS

Editors: Po-Ling Loh and Maxim Raginsky

Abstract

We present an asymptotically optimal (ϵ, δ) differentially private mechanism for answering multiple, adaptively asked, Δ-sensitive queries, settling the conjecture of Steinke and Ullman [2020]. Our algorithm has a significant advantage that it adds independent bounded noise to each query, thus providing an absolute error bound. Additionally, we apply our algorithm in adaptive data analysis, obtaining an improved guarantee for answering multiple queries regarding some underlying distribution using a finite sample. Numerical computations show that the bounded-noise mechanism outperforms the Gaussian mechanism in many standard settings.

1. Introduction

Differential privacy provides a framework to publish statistics of datasets that contain users' information, while preserving their privacy. Here, one assumes an underlying dataset $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}$ where x_i contains private information of user i. An analyst, which does not have access to the dataset, requests some statistics of the data. These statistics are provided by the dataset holder, however, by analyzing them, the analyst should not learn significant information on any specific datapoint x_i. We follow the standard framework of (ϵ, δ) *differential privacy* Dwork et al. (2006b,a) where the parameter ϵ quantifies the typical level of privacy and δ is (intuitively) the probability that the algorithm fails to preserve privacy (formally defined in Section 3).

Perhaps the most well-studied problem in differential privacy is *answering multiple queries*. The adaptive version is described by an interactive game between the dataset holder and the analyst: in each iteration $i = 1, \ldots, k$, the analyst submits a query $q_i \colon \mathcal{X}^n \to \mathbb{R}$. Then, the dataset holder should provide an approximate answer $a_i \approx q_i(\mathbf{x})$. Providing the exact answer may cause a leakage of private information and a common approach is to output $a_i = q_i(\mathbf{x}) + \eta_i$, where η_i is a random noise, whose outcome is unknown to the analyst. The goal is to keep the magnitude of noise as low as possible while preserving privacy.

Clearly, it is impossible to preserve privacy while answering arbitrary queries accurately. This happens when particular datapoints have significant influence over the outcome of the query: Here, an accurate answer to the query would necessarily leak information on these datapoints. To avoid this issue, it is common to assume that each datapoint can change the outcome by at most Δ. Formally, we use the standard notion of Δ-sensitive queries: For any user i and any datasets $\mathbf{x}$ and $\mathbf{x}'$ that differ only on entry i, we assume that $|q_i(\mathbf{x}) - q_i(\mathbf{x}')| \leq \Delta$. For example, q_i can be the average of some bounded statistic $h \colon \mathcal{X} \to [0, 1]$, where $q_i(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} h(x_i)$ and $\Delta = 1/n$.

Despite being a central problem in differential privacy, it was unknown what is the least amount of noise that should be added. This question can be formalized as follows:

© 2022 Y. Dagan & G. Kur.

Question 1 *Fix parameters ϵ, δ, k and Δ. What is the minimal noise-level α, such that there is an (ϵ, δ) differentially private algorithm that answers k Δ-sensitive queries with error at most α? (namely, $\forall i = 1 \cdots k, |a_i - q_i(\mathbf{x})| \leq \alpha$)*

One of the earliest algorithms, the *Gaussian mechanism* Dwork et al. (2006a), consists of adding independent Gaussian noise of standard deviation $\sigma = O(R)$, where $R := \Delta\sqrt{k \log 1/\delta}/\epsilon$. Namely, $a_i = q_i(\mathbf{x}) + \eta_i$ where $\eta_i \sim N(0, \sigma^2)$ are independent Gaussians. This yields a high-probability error of

$$\alpha = \max_{i=1,\ldots,k} |q_i(\mathbf{x}) - a_i| = \max_{i=1,\ldots,k} |\eta_i| \leq O(R\sqrt{\log k}),$$

since the maximum of k Gaussian random variables with standard deviation σ is bounded by $O(\sigma\sqrt{\log k})$ with high probability. This was the best known algorithm until long after, when Steinke and Ullman (2016) showed how to obtain an improved bound $\alpha \leq O(R\sqrt{\log \log k})$, by applying the same Gaussian mechanism and adding a smart algorithmic step that truncates the most-erroneous answers (via the *sparse vectors* algorithm Dwork et al. (2009); Hardt and Rothblum (2010); Dwork et al. (2010); Roth and Roughgarden (2010)). In the same paper, they also showed a lower bound of $\alpha \geq \Omega(R)$, for any $\delta \geq e^{-k}$. Later, Ganesh and Zhao (2020) showed how to obtain an improved bound of $O(R\sqrt{\log \log \log k})$ by replacing the Gaussian distribution with a generalized Gaussian and applying the same truncation technique of Steinke and Ullman (2016). We further note that for all $\delta \leq e^{-k}$, the optimal noise level is different, and equals $\Theta(k/\epsilon)$, using an algorithm that is in fact $(\epsilon, 0)$ private Hardt and Talwar (2010). Yet, it remained open whether one can match the lower bound of Steinke and Ullman (2016) and achieve a noise of $\alpha = O(R)$ for $\delta \geq e^{-o(k)}$. This was raised as an open problem by Steinke and Ullman (2020).

One feature common to these algorithms is that they rely on adding *unbounded noise*, and then, possibly, making a correction. Such an approach has multiple obvious disadvantages: (1) All the above-discussed algorithms fail to give a definite bound on the error that holds with probability 1; (2) The correction step (i.e. the sparse vector technique), if used, complicates the algorithm and (3) The numerical constants associated with the noise may significantly degrade if one uses correction techniques.

To guarantee a bounded noise, various prior works Liu (2018); Holohan et al. (2020) suggested to truncate known noise distributions such as the Gaussian and Laplace. Yet, this yields suboptimal algorithms and it is possible that specifically tailored bounded-noise distributions would provide better results. This gives rise to the following question:

Question 2 *What are the best mechanisms that rely on adding i.i.d. bounded noise? Can they provide the asymptotically optimal noise rate? Can they yield a reduced noise in practical settings?*

1.1. Main Results

In this paper, we provide a positive answer to the above two questions:

Theorem 1 *Let $k, n \in \mathbb{N}$, $\epsilon \in (0, 1]$, $\delta \in [e^{-k/\log^2(k)\log^4\log(k)}, 1/2]$ and $\Delta > 0$. There exists an algorithm for answering k adaptive Δ-sensitive queries that is (ϵ, δ) differentially private and further, its error is bounded as follows:*

$$\max_{i=1,\ldots,k} |a_i - q_i(\mathbf{x})| \leq O(R) := O(\Delta\sqrt{k \log(1/\delta)}/\epsilon), \quad \textit{with probability } 1.$$

Further, this is attained using an algorithm that adds i.i.d. noise of bounded magnitude to each answer.

In addition to providing an optimal error, this algorithm has additional benefits:

- The bound on the maximal error, $\max_i |a_i - q_i(\mathbf{x})| = O(R)$, holds with probabilty 1! This provides to the analyst definite bounds on the true answer $q_i(\mathbf{x})$, which is significantly more convenient in some settings. In comparison, the previous algorithms discussed above only guarantee a high probability error bound, which degrades at least as fast as $R\sqrt{\log 1/\beta}$ for confidence level $1-\beta$.
- The algorithm is simple: the noise added to each query is drawn i.i.d. from some simple closed-form density. This is compared to the previous algorithms discussed above that relied on an additional algorithmic truncation step.
- It yields better bounds than the Gaussian mechanism in many practical settings. This can be shown using an algorithm that computes upper bounds on the optimal noise level that is required to achieve (ϵ, δ) privacy (see Section 5; code available online).

We recall that our algorithm achieves an optimal noise only for $\delta \geq e^{-k/\log^2 k \log^4 \log k}$. A subsequent (which was essentially concurrent) independent work of Ghazi et al. (2020) provided an optimal rate for all $\delta \geq e^{-k}$, thus closing the gap between the upper and lower bounds that was left open for δ slightly larger than e^{-k}. Their algorithm smartly consists of permuting the queries and applying multiple stages of the sparse vector algorithm to reduce the noise. The advantages of our algorithm include the three items discussed above and the fact that it can answer adaptively asked queries, which also makes it applicable to adaptive data analysis.

Lastly, we argue that it is impossible to achieve an optimal bound using an algorithm that adds i.i.d. bounded noise for all $\delta \geq e^{-k}$: (Proof of Section C)

Theorem 2 (informal) *There is no (ϵ, δ) differentially private algorithm that adds bounded i.i.d. noise, that is asymptotically optimal in the regime $\delta \geq e^{-\omega(k/\log^2 k)}$, where $\omega()$ denotes a strict asymptotic inequality (we assume in the proof that the noise density is unimodal, yet we believe that this assumption is redundant).*

1.2. Application to adaptive data analysis.

Adaptive data analysis concerns of answering multiple adaptively asked queries on some underlying distribution P over a domain $\mathcal{X}$, while having access only to a finite i.i.d. sample $x_1, \ldots, x_n \sim P$ (Dwork et al., 2015; Hardt and Ullman, 2014). This scenario is common in statistics and machine learning, where an adaptive procedure or an algorithm are used to infer or learn properties of the distribution.

The standard setting can be formulated as an interaction between the dataset holder, that has access to n i.i.d. samples from P, and a statistical analyst whose goal is to infer properties of the distribution. In each iteration $t = 1, \ldots, k$ the algorithm submits a *statistical query* $q_i \colon \mathcal{X} \to [0, 1]$, and the goal of the dataset holder is to send an answer a_i that approximates the expectation $q_i(P) = \mathbb{E}_{x\sim P}[q_i(x)]$. The queries are asked adaptively, namely, q_i can depend on the previous answers $a_1, \ldots, a_{i-1}$. The goal is to answer all the queries with low error, ensuring that with probability at least $1-\beta$, $|a_i - q_i(P)| \leq \alpha$ for all i.

The straightforward approach is to answer each query q_i using the sample average, outputting $a_i = \frac{1}{n}\sum_{j=1}^{n} q_i(x_j)$. This gives a valid result with high probability for non adaptively-asked queries, namely, if $q_1, \ldots, q_n$ are given a priori. However, in case that they are asked adaptively, it is possible, and even likely in some scenarios, that they fit or adjust to the specifically-drawn sample. In such cases, the sample-mean will not provide a valid approximation to the true expectation $q_i(P)$. A solution suggested by Dwork et al. (2015) is to use a differentially private algorithm to answer the queries. Intuitively, this prevents the queries q_i from fitting to the data, since the previous answers $a_1 \cdots a_{i-1}$ are differentially private with respect to it. In particular Dwork et al. (2015) proved a *transfer theorem* that yields guarantees for adaptive data analysis given guarantees for the underlying differentially private algorithm. This was further improved by various authors: Bassily et al. (2021) gave asymptotically tight bounds, whereas Feldman and Steinke (2017, 2018); Rogers et al. (2020) proved improved instance-specific bounds. Finally, Jung et al. (2021) provided a new proof with significantly improved constants. Applying the transfer theorem of Jung et al. (2021) on the best known differentially private algorithms, one obtains the following guarantee: for all $\alpha, \beta \in (0, 1/2)$, there is a sample size

$$n(\alpha, \beta) = O\left(\min\left(\frac{\sqrt{k \log k \log^2(1/\alpha\beta)}}{\alpha^2}, \frac{\sqrt{k \log\log k \log^3(1/\alpha\beta)}}{\alpha^2}\right)\right), \tag{1}$$

and an algorithm that receives $n(\alpha, \beta)$ samples from some arbitrary distribution P, and answers k adaptively asked queries, with a high-probability error bound of

$$\Pr\left[\forall i = 1 \cdots k,\ |q_i(P) - a_i| \le \alpha\right] \ge 1 - \beta.$$

Here, the first argument in the right hand side of Eq. (1) corresponds to the Gaussian mechanism and the second to the algorithm of Steinke and Ullman (2016). In this paper, we obtain the following improved bound: (Proof in Section D)

Corollary 3 *For every $k \in \mathbb{N}$ and $\alpha, \beta \in (0, 1/2)$ such that $\alpha\beta \ge 4e^{-k/\log^2 k \log\log^4 k}$, there exists an algorithm for answering k adaptive statistical queries $q_i \colon \mathcal{X} \to [0, 1]$, with a sample size of*

$$n = O\left(\frac{\sqrt{k \log(1/\alpha\beta)}}{\alpha^2}\right),$$

that satisfies $\Pr\left[\forall i = 1 \cdots k,\ |q_i(P) - a_i| \le \alpha\right] \ge 1 - \beta$. More generally, this bound is also valid for answering 1-sensitive queries.[1]

This yields an optimal dependence both on k and β (Bassily et al., 2021), while removing logarithmic factors in k, α, β. Further, this algorithm can answers approximately twice as many queries as the Gaussian mechanism in a standard setting (see Section 5).

2. The abstract theorem

We present an abstract statement that provides guarantees for bounded-noise distributions, assuming that they satisfy some differential inequalities. We use the following notation for bounded-noise mechanisms:

1. When answering a Δ-sensitive query $q \colon \mathcal{X}^n \to \mathbb{R}$, the goal is to provide an approximation to the expected value of the query taken over a random dataset, $q(P^n) := \mathbb{E}_{\mathbf{x} \sim P^n}[q(\mathbf{x})]$.

Definition 4 *Given a function $f\colon (-1,1) \to (0,\infty)$, consider the continuous distribution μ_f with density*

$$\mu_f(\eta) = \frac{\exp(-f(\eta))}{Z_f}, \quad \textit{where } Z_f = \int_{-1}^{1} \exp^{-f(\eta)}\, d\eta.$$

Further, for any $R > 0$ denote by $\mu_{f,R}$ the scaling of μ_f by R, namely $\eta \sim \mu_{f,R}$ is obtained from sampling $\eta' \sim \mu_f$ and setting $\eta = R\eta'$. Equivalent, $\mu_{f,R}$ has density

$$\mu_{f,R}(\eta) = \frac{\exp(-f(\eta/R))}{Z_{f,R}}, \quad \textit{where } Z_{f,R} = \int_{-R}^{R} \exp^{-f(\eta/R)}\, d\eta = RZ_f.$$

Define by $M_{f,R}$ the mechanism that adds to each query a noise drawn independently from $\mu_{f,R}$.

Next, we present our abstract theorem that shows that some noise mechanisms are optimal, if f satisfies some desired properties. There are two essential properties: (1) μ_f decays to zero in the neighborhoods of -1 and 1, or, equivalently, $f(\eta) \to \infty$ as $\eta \to \pm 1$; and (2) μ does not decay too fast, which amounts to requiring that $|f'(\eta)|$ is bounded in terms of $f(\eta)$. In particular, we would like that $I(|f'(\eta)|) \leq f(\eta)$ for a monotonic increasing function I defined below. Any function f that satisfies these assumptions (and a couple more technical assumptions), yields DP mechanisms with asymptotically optimal error, for any $\delta \geq \delta_k^*$. The threshold δ_k^* improves (i.e. decreases) as the upper bound on $|f'(\eta)|$ improves, or, equivalently, as I increases.

Formally, let $I\colon [0,\infty) \to [0,\infty)$ be a continuous function that satisfies the following properties:

- $I(t) \leq t$ and $I(t) \geq c\sqrt{t}$ for any $t \geq C$, where $C, c > 0$ can be any constants independent of t.
- $I(t)$ is increasing in t and $I(t)/t$ is decreasing in t.

For example, $I(t) = t^\alpha$ for some $\alpha \in [1/2, 1]$ or $I(t) = t/\log^\alpha t$ for $\alpha \geq 0$. Now, we state some requirements on the function f that appears in the definition above:

1. f is symmetric, i.e. $f(-\eta) = f(\eta)$;
2. f diverges: $\lim_{\eta \to 1^-} f(\eta) = \lim_{\eta \to -1^+} f(\eta) = \infty$.
3. Bounded first derivative: $I(|f'(\eta)|) \leq f(\eta)$.
4. Bounded second derivative: $|f''(\eta)| \leq Cf(\eta)^2$, where $C > 0$ can be any constant independent of η.

Lastly, we define δ_k^*. For this purpose, define t^* as the unique solution to $t = kI(t)/2t$. Notice that such a unique solution exists as $I(t)/t$ is continuous and decreasing in t. Then, we define $\delta_k^* = e^{-I(t^*)/C_f}$ where $C_f > 0$ is a constant depending only on f. This yields the following theorem: (Proof in Section A)

Theorem 5

Let $I(t)$ and f satisfy the conditions above, let $k \in \mathbb{N}$ and δ_k^* be defined as above. Let $\Delta > 0$, $\epsilon \in (0,1]$, $\delta \in [\delta_k^*, 1/2]$ and define $R = C_f \Delta \sqrt{k \log 1/\delta}/\epsilon$ for some constant C_f depending only on f. Then, the mechanism $M_{f,R}$ is (ϵ, δ)-differentially private for answering k adaptive Δ-sensitive queries.

As a corollary, we obtain the following guarantees for specific functions f: (Proof in Section B)

Corollary 6 *The following functions f yield mechanisms $M_{f,R}$ with an optimal value of $R = \Theta(\Delta\sqrt{k\log(1/\delta)})/\epsilon)$, for any $\delta \geq \delta_k^*$:*

- *The function $f(\eta) = 1/(1-\eta^2)^p$ with $\delta_k^* = \exp\left(-C(p)k^{p/(p+2)}\right)$, for any $p \geq 2$, where $C(p)$ depends only on p.*
- *The function $f(\eta) = \exp\left(\exp\left(1/(1-\eta^2)\right)\right)$ with $\delta_k^* = \exp(-Ck/(\log^2 k \log^4 \log k))$, for some $C > 0$.*

3. Preliminaries

Neighboring datasets and Δ-sensitive queries. Given a domain $\mathcal{X}$ and $n \in \mathbb{N}$, a *dataset* is any element of $\mathcal{X}^n$. Two datasets $\mathbf{x}$ and $\mathbf{x}'$ are called *neighbors* if $\mathbf{x}$ and $\mathbf{x}'$ differ on exactly one entry. Given $\Delta > 0$, a Δ-*sensitive query* is any function $q\colon \mathcal{X}^n \to \mathbb{R}$ such that for any two neighboring datasets $\mathbf{x}$ and $\mathbf{x}'$, it holds that $|q(\mathbf{x}) - q(\mathbf{x}')| \leq \Delta$.

Interactive and non-interactive query-answering. The interactive setting can be viewed as an interactive game between two parties: (1) a dataset holder, which has access to some dataset $\mathbf{x} \in \mathcal{X}^n$, and (2) an analyst that has no information on $\mathbf{x}$. In each iteration $t = 1, \ldots, k$, the analyst submits to the dataset-holder some query $q_i\colon \mathcal{X}^n \to \mathbb{R}$, who replies with an answer a_i, that approximates the value $q_i(\mathbf{x})$. Notice that here q_i can depend only on the previous answers $a_1, \ldots, a_{i-1}$. In comparison, in the non-interactive setting the analyst submits all the queries ahead of time, and then the dataset holder answers them.

Differential privacy. Fix some mechanism M for answering k Δ-sensitive queries and fix $\epsilon, \delta > 0$. Let A denote any query-asking strategy of the analyst that defines each query q_i as a function of the previous answers $a_1, \ldots, a_{i-1}$. We say that M satisfies (ϵ, δ)-differential privacy if for any two neighboring datasets $\mathbf{x}$ and $\mathbf{x}'$, any analyst A and any subset $U \subseteq \mathbb{R}^k$ of possible answers,

$$\Pr[(a_1, \ldots, a_k) \in U \mid \mathbf{x}, A] \leq e^{\epsilon} \Pr[(a_1, \ldots, a_k) \in U \mid \mathbf{x}', A] + \delta.$$

Intuitively, the distributions over the answers given any two neighboring datasets are similar.

4. Proof Sketch

We provide a proof sketch for Theorem 5, assuming that $\Delta = 1$. To simplify the presentation, we assume that the queries $q_1, \ldots, q_n$ are fixed and non-adaptive. The proof consists of two steps: first, we reduce the problem to showing a concentration inequality on a sum of independent variables, and secondly, we bound this sum.

REDUCING TO A CONCENTRATION INEQUALITY

In this section, our goal is to show that it suffices to prove Eq. (5) ahead, which corresponds to bounding a weighted sum of $f'(\eta_1), \ldots, f'(\eta_k)$, for randomly drawn $\eta_1, \ldots, \eta_k \overset{\text{i.i.d.}}{\sim} \mu_f$. Denote by $\boldsymbol{q} := (q_1, \ldots, q_k)$ and $\boldsymbol{a} := (a_1, \ldots, a_k)$ the vectors of queries and answers, respectively, and let $\Pr_{\cdot|\mathbf{x}}$ and $\text{density}_{\cdot|\mathbf{x}}$ denote the conditional probability and density of $\boldsymbol{a}$ given the dataset $\mathbf{x}$,

respectively. Recall that we want to show that for any two neighboring datasets $\mathbf{x}$ and $\mathbf{y}$ and any subset $U \subseteq \mathbb{R}^k$, we have that $\Pr_{\cdot|\mathbf{x}}[\boldsymbol{a} \in U] \leq e^{\epsilon} \Pr_{\cdot|\mathbf{y}}[\boldsymbol{a} \in U] + \delta$. A simple argument shows that it suffices to prove that

$$\Pr_{\cdot|\mathbf{x}}\left[\frac{\text{density}_{\cdot|\mathbf{x}}[\boldsymbol{a}]}{\text{density}_{\cdot|\mathbf{y}}[\boldsymbol{a}]} \geq e^{\epsilon}\right] \leq \delta \ . \tag{2}$$

Intuitively, this means that only a small fraction of the possible answers $\boldsymbol{a} \in \mathbb{R}^k$ are significantly more likely given $\mathbf{x}$ compared to $\mathbf{y}$. Recall that the answers a_i are obtained by adding an i.i.d. noise whose density equals $\text{noise}(\eta_i) := \exp(-f(\eta_i/R))/Z_{f,R}$, therefore

$$\text{density}_{\cdot|\mathbf{x}}[\boldsymbol{a}] = \prod_{i=1}^{k} \text{noise}(a_i - q_i(\mathbf{x})) = \prod_{i=1}^{k} \exp\left(-f\left(\frac{a_i - q_i(\mathbf{x})}{R}\right)\right)/Z_{f,R} \ .$$

Substituting this in Eq. (2) and taking a log inside the $\Pr[]$, one obtains

$$\Pr_{\cdot|\mathbf{x}}\left[-\sum_{i=1}^{k} f\left(\frac{a_i - q_i(\mathbf{x})}{R}\right) + \sum_{i=1}^{k} f\left(\frac{a_i - q_i(\mathbf{y})}{R}\right) \geq \epsilon\right] \leq \delta. \tag{3}$$

We substitute $\eta_i = (a_i - q_i(\mathbf{x}))/R$ and $v_i = (q_i(\mathbf{x}) - q_i(\mathbf{y}))/R$, which also implies that $(a_i - q_i(\mathbf{y}))/R = \eta_i + v_i$. Notice that $\eta_i \sim \mu_{f,1} = \mu_f$, namely η_i is drawn from the normalized noise supported in $(-1, 1)$ and notice that $v_i \in [-1/R, 1/R]$, since q_i is 1-sensitive. Then, Eq. (3) translates to

$$\Pr_{\boldsymbol{\eta}\sim\mu_f^k}\left[\sum_{i=1}^{k} f(\eta_i + v_i) - \sum_{i=1}^{k} f(\eta_i) \geq \epsilon\right] \leq \delta. \tag{4}$$

We then use the second-degree Taylor expansion to obtain $f(\eta_i+v_i) = f(\eta_i)+v_i f'(\eta_i)+v_i^2 f''(\xi_i)/2$ for some ξ_i in the line connecting η_i and $\eta_i + v_i$, and particularly, $\xi_i \in [\eta_i - 1/R, \eta_i + 1/R]$. Substituting this in Eq. (4) and substituting $v_i = u_i/R$, it suffices to prove the second inequality below:

$$\Pr_{\boldsymbol{\eta}\sim\mu_f^k}\left[\sum_{i=1}^{k} v_i f'(\eta_i) + \sum_{i=1}^{k} v_i^2 f''(\xi_i) \geq \epsilon\right] \leq \Pr_{\boldsymbol{\eta}\sim\mu_f^k}\left[\frac{1}{R}\sum_{i=1}^{k} u_i f'(\eta_i) + \frac{1}{2R^2}\sum_{i=1}^{k} \max_{\xi_i} |f''(\xi_i)| \geq \epsilon\right] \leq \delta \tag{5}$$

where $u_i \in [-1, 1]$, we substituted $v_i^2 = u_i^2/R^2 \leq 1/R^2$ and the maximum is taken over $\xi_i \in [\eta_i - 1/R, \eta_i + 1/R]$. Notice that it is possible that $\xi_i \notin (-1, 1)$, and for these values, we use the convention $f''(\xi_i) = \infty$. We will bound separately by $\epsilon/2$ the sums that correspond to the first and the second derivatives.

PROVING THE CONCENTRATION INEQUALITY.

Before sketching the actual concentration inequality that is used to bound the sum of first derivatives, we give an intuition by applying a central limit theorem, which is valid for any fixed δ as $k \to \infty$.

Central limit theorem for the sum of first-derivatives. Here, we assume for simplicity that $u_i \in \{-1, 1\}$, which implies that $\mathrm{Var}(u_i f'(\eta_i)) = \mathrm{Var}(f'(\eta_i)) := \sigma^2$. Further, notice that since f is assumed to be symmetric, we have that $\mathbb{E}[u_i f'(\eta_i)] = u_i \mathbb{E}[f'(\eta_i)] = 0$ for all i. Thus, for any $t \geq 0$,

$$\lim_{k\to\infty} \Pr\left[\frac{\sum_{i=1}^k u_i f'(\eta_i)}{\sqrt{k}\sigma} > t\right] = \int_t^\infty \frac{e^{-s^2/2}}{\sqrt{2\pi}} ds \leq e^{-t^2/2}\,; \quad \text{where } \sigma^2 = \mathrm{Var}(f'(\eta_i)) \ .$$

If we fix $\delta > 0$, take $t = \sqrt{\log(2/\delta)}$ and $R = 2\sigma\sqrt{k\log(2/\delta)}/\epsilon$, we obtain that for a sufficiently large k,

$$\Pr\left[\frac{\sum_{i=1}^k u_i f'(\eta_i)}{R} > \epsilon/2\right] \leq \delta/2.$$

This is what we wanted to prove, in terms of the sum over first derivatives, and if we prove a similar statement with respect to the sum of second derivatives, the proof concludes. Yet, this bound holds for any *fixed* δ in the limit $k \to \infty$. Instead, we want a bound that holds when $k \to \infty$ and $\delta \to 0$ *simultaneously*.

Non-asymptotic bound for the sum of first derivatives. Here, we would like to prove a non-asymptotic result. The standard approach to bounding a sum of independent random variables, $\sum_{i=1}^k X_i$, is to prove that each individual variable X_i concentrates, and this should imply a concentration inequality for the sum. Perhaps the most well-known concentration inequality is Chernoff-Hoeffding, which assumes that the variables X_i are bounded. Other inequalities assume that the X_is have a bounded tail. For example, Bernstein's inequality is valid if there exists some constant $C > 0$ such that

$$\Pr[|X_i| > t] \leq C\exp(-t/C). \tag{6}$$

In our case, substituting $X_i = u_i f'(\eta_i)$, we cannot guarantee such behavior. Instead, we can guarantee

$$\Pr[|X_i| > t] = \Pr[|u_i f'(\eta_i)| > t] \leq C\exp(-I(t)), \tag{7}$$

where $I(t) \ll t$ is the function given in the theorem statement. Next, we describe how to obtain Eq. (7) and then, we explain how to bound the sum $\sum_i u_i f'(\eta_i)$ assuming Eq. (7).

To prove Eq. (7), one can use the assumption $f(\eta_i) \geq I(|f'(\eta_i)|)$ and the fact that $I(t)$ is monotonic non-decreasing, and integrate:

$$\begin{aligned}\Pr[|u_i f'(\eta_i)| > t] \leq \Pr[|f'(\eta_i)| > t] &= \frac{1}{Z_f}\int_{\substack{\eta\in(-1,1):\\ |f'(\eta)|\geq t}} e^{-f(\eta)}d\eta \leq \frac{1}{Z_f}\int_{\substack{\eta\in(-1,1):\\ |f'(\eta)|\geq t}} e^{-I(|f'(\eta)|)}d\eta \\ &\leq \frac{1}{Z_f}\int_{\substack{\eta\in(-1,1):\\ |f'(\eta)|\geq t}} e^{-I(t)}d\eta \leq \frac{2}{Z_f}e^{-I(t)}.\end{aligned}$$

Next, we explain how to bound the sum of independent variables satisfying Eq. (7). Here we go along the lines of Bakhshizadeh et al. (2020) which uses the known idea of *truncation*, as explained below. We start by explaining the standard approach that is used for bounded random variables or variables satisfying Eq. (6), and then explain how to adapt these ideas to our setting. The standard approach is via an analysis of the moment generating function: for any $\theta > 0$, we can compute

$$\mathbb{E}\left[\exp\left(\theta\sum_{i=1}^k X_i\right)\right] = \prod_{i=1}^k \mathbb{E}\left[\exp(\theta X_i)\right],$$

and use Markov's inequality to bound:

$$\Pr\left[\sum_i X_i > t\right] = \Pr\left[\exp\left(\theta \sum_i X_i\right) > \exp(\theta t)\right] \leq \frac{\mathbb{E}\left[\exp\left(\theta \sum_i X_i\right)\right]}{\exp(\theta t)} = \frac{\prod_{i=1}^k \mathbb{E}\left[\exp\left(\theta X_i\right)\right]}{\exp(\theta t)} .$$

We can now optimize over $\theta > 0$ to obtain the known inequalities.

Next, we move to our setting, substituting $X_i = u_i f'(\eta_i)$. Since $I(t) \ll t$ and Eq. (6) does not hold, we cannot use the MGF bound, because $\mathbb{E}[\exp(\theta X_i)] = \infty$ for all $\theta > 0$. This is called the *heavy-tailed* regime. A standard approach is to truncate the random variables. Given some fixed $L \geq 0$, we define $X_i^{\leq L} = X_i \mathbb{1}(X_i \leq L)$ and notice that $X_i^{\leq L}$ is bounded, hence its MGF is finite for any $\theta > 0$. Then, in order to bound the probability that $\sum_i X_i > t$, we first bound the probability that $\sum_i X_i^{\leq L} > t$ and then bound the probability that there exists an i such that $X_i > L$, leading to the following bound:

$$\Pr\left[\sum_{i=1}^k X_i > t\right] \leq \Pr\left[\sum_{i=1}^k X_i^{\leq L} > t\right] + \sum_{i=1}^k \Pr[X_i > L].$$

The first term can be bounded using the moment generating function, and the second term is simply bounded by $Cke^{-I(L)}$, using Eq. (7). Optimizing over the parameter θ in the MGF bound and over the truncation parameter L, one obtains the following bound, assuming that the X_is have zero mean:

$$\Pr\left[\sum_{i=1}^k X_i > t\right] \leq 2e^{-t^2/kC'} + C'ke^{-I(t^*)/C'} \tag{8}$$

for all $t \leq t^*$ where t^* is the solution to $t = kI(t)/2t$. Here, t^* is the same parameter as defined in Theorem 5 and $C' > 0$ possibly depends on $I(t)$. We note that the first term is the analogue of the CLT, and it dominates the second term for $t \leq t^*$, hence, we obtain the desired bound.

Bounding the sum of second derivatives. Recall that we want to show that

$$\Pr\left[\frac{1}{2R^2}\sum_{i=1}^k \max_{\xi_i \in [\eta_i - 1/R, \eta_i + 1/R]} |f''(\xi_i)| \leq \epsilon/2\right] \geq 1 - \delta/2. \tag{9}$$

First, using the condition that $|f''(\eta)| \leq f(\eta)^2$, it suffices to show that

$$\Pr\left[\frac{1}{2R^2}\sum_{i=1}^k \max_{\xi_i} f(\xi_i)^2 \leq \epsilon/2\right] \geq 1 - \delta/2. \tag{10}$$

Next, we show that with high probability over η_i, we have that $f(\xi_i) \leq 2f(\eta_i)$. Since the density is proportional to $\exp(-f(\eta_i))$, $f(\eta_i)$ is small with high probability, so it is sufficient to show that if $f(\eta_i)$ is not too large, then $f(\eta_i \pm 1/R) \leq 2f(\eta_i)$. For that purpose, we use the condition that $|f'(\eta_i)| \leq f(\eta_i)^2$, which guarantees that if f is not very large, then it cannot grow very fast. This will conclude that with high probability $f(\xi_i) \leq 2f(\eta_i)$, and by taking a union bound, this holds with high probability simultaneously for all i. Then, Eq. (10) translates to

$$\Pr\left[\frac{1}{2R^2}\sum_{i=1}^k (2f(\eta_i))^2 \leq \epsilon/2\right] \geq 1 - \delta/2.$$

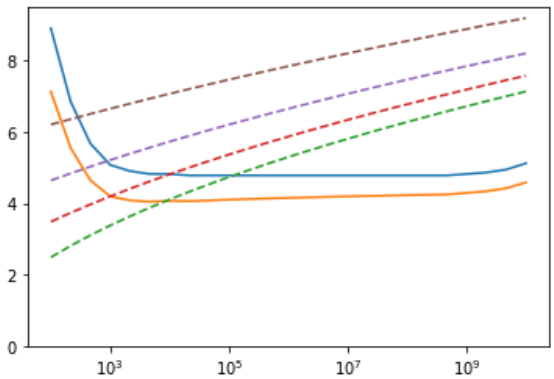
Similarly to the arguments above regarding the first derivative, we can show that $\Pr[f(\eta_i)^2 > t] \leq C\exp(-\sqrt{t})$. Again, we use a concentration inequality similar to Eq. (8) to bound this sum. Following Eq. (5) and the bound on the sum of first derivatives, this concludes the proof.

5. Simulations

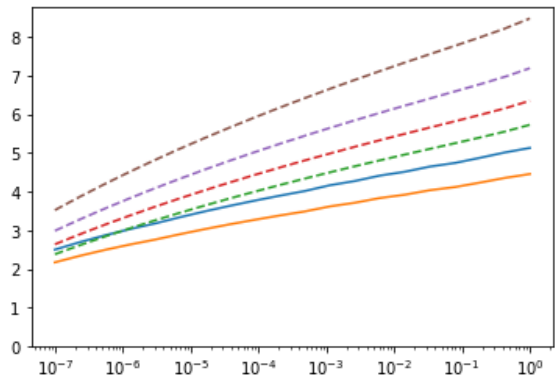
We compare the numerical noise-levels of bounded-noise mechanisms to the Gaussian mechanism, for fixed values of k, ϵ and δ. We used a computer program to derive tighter noise bounds than the ones appearing in the proof, by an exact computation of a suitable moment generating function (the formal derivation appears in Section E). We note that similar techniques can be used to obtain bounds on any mechanism that uses i.i.d. noise. Yet, for the Gaussian mechanism we used the exact optimal noise level, computed in Balle and Wang (2018). Still, the upper bounds on the bounded-noise mechanism outperform those optimally computed values for the Gaussian mechanism. We did not compare to the other mechanisms that have better asymptotic noise than the Gaussian mechanism, as the constants associated with their bounds are significantly worse.

The comparison appears in Figure 1. For the bounded mechanism, we plot both the absolute bound on the noise and the 0.95 probability bound on the maximal noise over k queries, whereas for the Gaussian mechanism we plotted high probability bounds on the maximal noise with different confidence levels. It is worth noting that the gap between the bounded noise and the Gaussian mechanism increases as k grows, as expected. For the fixed setting of $\epsilon = 0.1$ and $\delta = 10^{-10}$, the 0.95-probability bound for the bounded-noise mechanism matches the 0.95 bound by the Gaussian mechanism already at $k = 10^3$, and it is 29% less at $k = 10^6$. Further, the absolute bound for the bounded-noise mechanism is lower by 28% than the 0.999-probability bound of the Gaussian mechanism at $k = 10^6$.

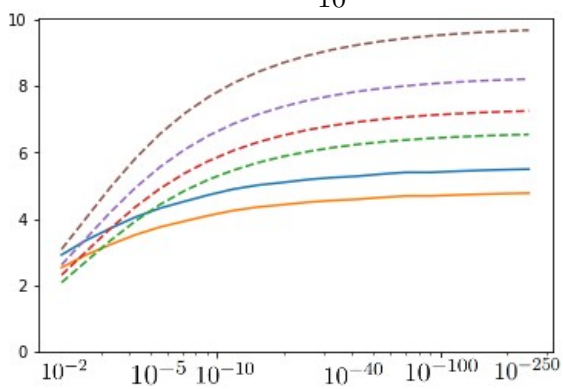
Further, we present numerical comparisons for adaptive data analysis in Figure 2. We used the same setting that was plotted in Jung et al. (2021): we set the values of $\alpha = 0.1$ and $\beta = 0.05$, and for multiple values of n, we computed the number of adaptive queries that can be answered while keeping all the errors below α with probability $1 - \beta$. Here, the bounded noise mechanism can answer at least twice many queries as the Gaussian mechanism for any $n \geq 8 \cdot 10^5$ and it significantly outperforms the Gaussian mechanism also for smaller values of k.



(*a*) k ranges from 100 to 10^{10}, $\epsilon = 0.1$ and $\delta = 10^{-10}$



(*b*) ϵ ranges from 10^{-7} to 1, $k = 10^6$ and $\delta = 10^{-10}$



(*c*) δ ranges from 10^{-2} to 10^{-250}, $k = 10^6$ and $\epsilon = 0.1$

Figure 1: The errors of different mechanisms are plotted as a function of k, ϵ and δ. The solid lines correspond to bounded noise mechanism with $f(\eta) = 1/(1-\eta^2)^2$. In all of the plots, the upper solid line corresponds to the absolute bound on the noise, while the lower solid line corresponds to a 0.95-probability bound on the maximal error over the k queries. The dashed lines corresponds to the Gaussian mechanism, and they correspond to bounds on the maximal error that hold with probabilities $1 - 10^{-6}, 0.999.0.95$ and 0.5 (larger noise corresponds to a higher probability). The values on the x-axis are described in each figure separately and the y-axis corresponds to the noise divided by $\sqrt{k \log(1/\delta)}/\epsilon$.

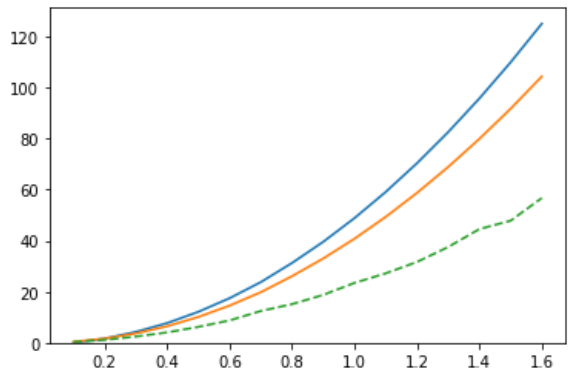


Figure 2: The number of queries k (in thousands) that can be answered as a function of n (in millions), while retaining $(\alpha = 0.1, \beta = 0.05)$-validity. The top line corresponds to the bounded noise mechanism with $f(\eta) = 1/(1-\eta^2)$, the middle line to $f(\eta) = 1/(1-\eta^2)^2$ (which is the same mechanism tested in Figure 1) and the dashed line corresponds to the Gaussian mechanism.

References

Milad Bakhshizadeh, Arian Maleki, and Victor H de la Pena. Sharp concentration results for heavy-tailed distributions. *arXiv preprint arXiv:2003.13819*, 2020.

Borja Balle and Yu-Xiang Wang. Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. In *International Conference on Machine Learning*, pages 394–403. PMLR, 2018.

Raef Bassily, Kobbi Nissim, Adam Smith, Thomas Steinke, Uri Stemmer, and Jonathan Ullman. Algorithmic stability for adaptive data analysis. *SIAM Journal on Computing*, (0):STOC16–377, 2021.

Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 486–503. Springer, 2006a.

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, 2006b.

Cynthia Dwork, Moni Naor, Omer Reingold, Guy N Rothblum, and Salil Vadhan. On the complexity of differentially private data release: efficient algorithms and hardness results. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 381–390, 2009.

Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N Rothblum. Differential privacy under continual observation. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 715–724, 2010.

Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Leon Roth. Preserving statistical validity in adaptive data analysis. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 117–126, 2015.

Vitaly Feldman and Thomas Steinke. Generalization for adaptively-chosen estimators via stable median. In *Conference on Learning Theory*, pages 728–757. PMLR, 2017.

Vitaly Feldman and Thomas Steinke. Calibrating noise to variance in adaptive data analysis. In *Conference On Learning Theory*, pages 535–544. PMLR, 2018.

Arun Ganesh and Jiazheng Zhao. Privately answering counting queries with generalized gaussian mechanisms. *arXiv preprint arXiv:2010.01457*, 2020.

Badih Ghazi, Ravi Kumar, and Pasin Manurangsi. On avoiding the union bound when answering multiple differentially private queries. *arXiv preprint arXiv:2012.09116*, 2020.

Moritz Hardt and Guy N Rothblum. A multiplicative weights mechanism for privacy-preserving data analysis. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 61–70. IEEE, 2010.

Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 705–714, 2010.

Moritz Hardt and Jonathan Ullman. Preventing false discovery in interactive data analysis is hard. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 454–463. IEEE, 2014.

Naoise Holohan, Spiros Antonatos, Stefano Braghin, and Pól Mac Aonghusa. The bounded laplace mechanism in differential privacy. *Journal of Privacy and Confidentiality*, 10(1), 2020.

Christopher Jung, Katrina Ligett, Seth Neel, Aaron Roth, Saeed Sharifi-Malvajerdi, and Moshe Shenfeld. A new analysis of differential privacy's generalization guarantees. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 9–9, 2021.

Fang Liu. Generalized gaussian mechanism for differential privacy. *IEEE Transactions on Knowledge and Data Engineering*, 31(4):747–756, 2018.

Ryan Rogers, Aaron Roth, Adam Smith, Nathan Srebro, Om Thakkar, and Blake Woodworth. Guaranteed validity for empirical approaches to adaptive data analysis. In *International Conference on Artificial Intelligence and Statistics*, pages 2830–2840. PMLR, 2020.

Aaron Roth and Tim Roughgarden. Interactive privacy via the median mechanism. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 765–774, 2010.

Thomas Steinke and Jonathan Ullman. Between pure and approximate differential privacy. *Journal of Privacy and Confidentiality*, 7(2):3–22, 2016.

Thomas Steinke and Jonathan Ullman. Open problem - avoiding the union bound for multiple queries, 2020. URL https://differentialprivacy.org/open-problem-avoid-union/.

Appendix A. Abstract upper bound: Proof of Theorem 5

First, notice that in the proof sketch we assume the queries to be non-adaptive. Hence, we start by explaining the differences that one has to make in order to adjust to the adaptive setting. Then, we proceed with the formal proof.

The adaptive vs. the non-adaptive setting. In the non-adaptive setting, the queries $q_1, \ldots, q_n$ are asked ahead of time, and assumed to be fixed. Notice that in Eq. (5), the queries q_i come into play via u_i. Indeed, recall that $u_i = Rv_i = q_i(\mathbf{x}) - q_i(\mathbf{y})$. In this setting, the u_i are fixed numbers. Since the noise entries η_i are i.i.d., we derive that $\sum_i u_i f'(\eta_i)$ is a sum of i.i.d. random variables. In order to bound them, we apply a concentration inequality for a sum of i.i.d. variables.

In comparison, in the adaptive setting, q_i is asked after observing the previous answers $a_1, \ldots, a_{i-1}$. Since q_i depends on $a_1, \ldots, a_{i-1}$, then u_i depends on $\eta_1, \ldots, \eta_{i-1}$. In particular, the summands in $u_i f'(\eta_i)$ are no longer i.i.d. Yet, since q_i is only a function of $\eta_1, \ldots, \eta_{i-1}$, then u_i is only a function of $\eta_1, \ldots, \eta_{i-1}$ and it is independent on $\eta_i, \ldots, \eta_k$. Since the η_i variables are i.i.d., it holds that $\mathbb{E}[u_i f'(\eta_i) \mid \eta_1, \ldots, \eta_{i-1}] = u_i \mathbb{E}[f'(\eta_i) \mid \eta_1, \ldots, \eta_{i-1}] = 0$. In particular, the partial sums of $\sum_i f'(\eta_i) u_i$ constitute of a Martingale whose deviation can be bounded using a concentration inequality.

FORMAL PROOF

Here, we use asymptotic notation, e.g. $O()$, to hide constants that might depend on the log-density function f. Notice that it suffices to prove for $\Delta = 1$ (we can always scale the the queries and the noise by the same amount, while retaining (ϵ, δ)-privacy).

We start with a simple sufficient condition for the mechanism to be (ϵ, δ) differentially private:

Lemma 7 *Let $\epsilon, \delta > 0$ and let $k \in \mathbb{N}$. Let f and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k) \overset{iid}{\sim} \mu_{f,1}$. Let $R > 0$ and assume that for any random variables $v_1, \ldots, v_n \in [-1/R, 1/R]$ such that v_i is a deterministic function of $\eta_1, \ldots, \eta_{i-1}$, it holds that*

$$\Pr\left[\sum_{j=1}^{k} f(\eta_j + v_j) \leq \sum_{j=1}^{k} f(\eta_j) + \epsilon\right] \geq 1 - \delta. \tag{11}$$

Then, $M_{f,R}$ is (ϵ, δ) differentially private.

Proof First, we can assume that $Z_{f,R} = \int_{-R}^{R} e^{-f(\eta/R)} d\eta = 1$ (otherwise, f can be replaced with $f + \log Z_{f,R}$). Let $\mathbf{x}$ and $\mathbf{x}'$ denote two neighboring datasets, let $\boldsymbol{a} = (a_1, \ldots, a_k)$ denote the random output of the algorithm on input $\mathbf{x} = (x_1, \ldots, x_n)$ denote by $\boldsymbol{a}' = (a'_1, \ldots, a'_k)$ its output on input $\mathbf{x}'$. Let $U \subseteq \mathbb{R}^k$ and our goal is to show that $\Pr[\boldsymbol{a} \in U] \leq e^{\epsilon} \Pr[\boldsymbol{a}' \in U] + \delta$. Denote $v_j = (q_j(\mathbf{x}) - q_j(\mathbf{x}'))/R$ for $j \in [k]$ and notice that v_j is a deterministic function of $\eta_1, \ldots, \eta_{j-1}$. Denote

$$G = \left\{ \boldsymbol{\eta} \in \mathbb{R}^k \colon \sum_j f(\eta_j) \geq \sum_j f(\eta_j + v_j) - \epsilon \right\} .$$

Notice that by Eq. (11), $\Pr[\boldsymbol{\eta} \notin G] \leq \delta$. Denote $(U - \boldsymbol{b})/c = \{(\boldsymbol{u} - \boldsymbol{b})/c \colon u \in U\}$ for any $\boldsymbol{b} \in \mathbb{R}^k$ and $c \neq 0$, define $\boldsymbol{q}(\mathbf{x}) = (q_1(\mathbf{x}), \ldots, q_k(\mathbf{x}))$, notice that $a_j = q_j(\mathbf{x}) + R\eta_j$ and estimate:

$$\begin{aligned}
\Pr[\boldsymbol{a} \in U] &= \Pr\left[R\boldsymbol{\eta} \in (U - \boldsymbol{q}(\mathbf{x}))\right] \leq \Pr\left[\boldsymbol{\eta} \in \left((U - \boldsymbol{q}(\mathbf{x}))/R\right) \cap G\right] + \Pr\left[\boldsymbol{\eta} \notin G\right] \\
&\leq \Pr\left[\boldsymbol{\eta} \in \left((U - \boldsymbol{q}(\mathbf{x}))/R\right) \cap G\right] + \delta = \int_{\boldsymbol{u} \in ((U - \boldsymbol{q}(\mathbf{x}))/R) \cap G} e^{-\sum_i f(u_i)} du + \delta \\
&\leq \int_{\boldsymbol{u} \in ((U - \boldsymbol{q}(\mathbf{x}))/R) \cap G} e^{\epsilon - \sum_i f(u_i + v_i)} du + \delta = e^{\epsilon} \Pr\left[\boldsymbol{\eta} \in \left((U - \boldsymbol{q}(\mathbf{x}) + R\boldsymbol{v})/R\right) \cap G\right] + \delta \\
&= e^{\epsilon} \Pr\left[\boldsymbol{\eta} \in \left(\left(U - \boldsymbol{q}(\mathbf{x}')\right)/R\right) \cap G\right] + \delta \leq e^{\epsilon} \Pr\left[R\boldsymbol{\eta} \in \left(U - \boldsymbol{q}(\mathbf{x}')\right)\right] + \delta \\
&= e^{\epsilon} \Pr[\mathbf{x}' \in U] + \delta.
\end{aligned}$$

■

To prove that Eq. (11) holds, one can approximate f by its second-degree Taylor expansion, thus deriving the following statement.

Lemma 8 *Let $\epsilon, \delta, R > 0$ and let $k \in \mathbb{N}$. Let f and let $\boldsymbol{\eta} \overset{\text{i.i.d.}}{\sim} \mu_{f,1}$. Assume that for any $u_1, \ldots, u_n \in [-1, 1]$ such that u_j is a deterministic function$\eta_1, \cdots, \eta_{j-1}$, the following holds:*

$$\Pr\left[\forall i,\ \eta_i \in (-1 + 1/R, 1 - 1/R),\ \textit{and}\ \left|\sum_{i=1}^{k} \frac{u_i f'(\eta_i)}{R}\right| + \sum_{i=1}^{k} \max_{\xi_i \colon |\xi_i - \eta_i| \leq 1/R} \frac{|f''(\xi_i)|}{2R^2} \leq \epsilon\right] \geq 1 - \delta. \tag{12}$$

Then, $M_{f,R}$ is (ϵ, δ) differentially private.

Proof We will show that Eq. (12) implies Eq. (11), where u_i in Eq. (12) is replaced with Rv_i in Eq. (11). In particular, notice that we can write $f(\eta_j + v_j)$ using the Taylor expansion $f(\eta_j + v_j) = f(\eta_j) + f'(\eta_j)v_j + f''(\xi_j)v_j^2/2$ where ξ_j is a point in the line connecting η_j and $\eta_j + v_j$. We have

$$\begin{aligned}
\sum_j f(\eta_j + v_j) &= \sum_j f(\eta_j) + \sum_j f'(\eta_j)v_j + \sum_j \frac{f''(\xi_j)v_j^2}{2} \\
&\leq \sum_j f(\eta_j) + \left|\sum_j \frac{f'(\eta_j)u_j}{R}\right| + \sum_j \max_{\xi_j \colon |\xi_j - \eta_j| \leq 1/R} \frac{|f''(\xi_j)|}{2R^2}.
\end{aligned}$$

Thus, whenever the high-probability event in Eq. (12) holds, the event of Eq. (11) also holds. In particular, Eq. (12) implies Eq. (11), which concludes the proof. ■

To apply Lemma 8, we would like to prove concentration of the sum of the derivatives of f. In order to analyze the concentration properties of a sum of random variables, it is common to consider each variable separately and then use a concentration result for sums. We start by providing a definition of what it means for a random variable to concentrate:

Definition 9 *Given $C > 0$ and a function $I(t) \colon [0, \infty) \to \mathbb{R}$, we say that a random variable X is $(I(t), C)$ bounded if for all t, $\Pr[|X| > t] \leq C \exp(-I(t))$.*

Given a sum of $(I(t), C)$ bounded variables, we can obtain the following concentration inequality, that is proven in Section A.1, using ideas from Bakhshizadeh et al. (2020).

Proposition 10 *Let $X_1 \cdots X_n$ be a Martingale, namely, for all i, $\mathbb{E}[X_i \mid X_1, \ldots, X_{i-1}] = 0$. Further, assume that there exists a function $I\colon [0,\infty) \to [0,\infty)$ and $C > 0$ such that*

$$\Pr[|X_i| > t \mid X_1 \cdots X_{i-1}] \leq C \exp(-I(t)),$$

and additionally, $I(t) \leq t$ and $I(t)/t$ is monotonic decreasing. Let $M > 0$ be such that $M \geq C \int_0^\infty (t^2 + 2t) e^{-I(t)/2}$ and let t^* be the unique solution to

$$t = MnI(t)/2t.$$

Then, for any $t > 0$,

$$\Pr\left[\left|\sum_{i=1}^n X_i\right| > t\right] \leq \begin{cases} 2e^{-t^2/2Mn} + Cne^{-I(t^*)} & t \leq t^* \\ 2e^{-I(t)/4} + Cne^{-I(t)} & t > t^* \end{cases}.$$

To give some intuition, we note that for any $t \leq t^*$, the first term dominates the second, and here, we get a sub-Gaussian concentration, namely, the tail behaves as a Gaussian tail, where the variance of the Gaussian is replaced with Mn, where M is just a constant that depends on $I(t)$. Yet, one could not hope for a sub-Gaussian concentration for all $t \geq 0$. After all, a single variable X_i decays slower than a Gaussian. At some point, the heavy tail of the single X_i will dominate the sub-Gaussian tail of the sum, and this happens exactly at t^*. From that point onward, the tail is dictated by the function $I(t)$.

In order to apply Proposition 10, we would like to show concentration properties of a single instance of $f'(\eta)$ and $f''(\eta)$. This follows from the fact that f' and f'' are bounded in terms of some function of f, using the following simple lemma:

Lemma 11 *Let η be a random variable supported on $(-1,1)$, with density $e^{-f(\eta)}/Z$ where Z is the normalizing constant and $f(\eta) > 0$. Assume that $h\colon (-1,1) \to \mathbb{R}$ is such that $f(\eta) \geq I(|h(\eta)|)$ for all η, for some increasing $I\colon (0,\infty) \to (0,\infty)$. Then, for any $t \geq 0$,*

$$\Pr[|h(\eta)| \geq t] \leq \frac{2}{Z} e^{-I(t)}.$$

Proof A simple calculation shows

$$\begin{aligned}\Pr[|h(\eta)| \geq t] = \frac{1}{Z} \int_{\substack{\eta \in (-1,1):\\ |h(\eta)| \geq t}} e^{-f(\eta)} d\eta &\leq \frac{1}{Z} \int_{\substack{\eta \in (-1,1):\\ |h(\eta)| \geq t}} e^{-I(|h(\eta)|)} d\eta \\ &\leq \frac{1}{Z} \int_{\substack{\eta \in (-1,1):\\ |h(\eta)| \geq t}} e^{-I(t)} d\eta \leq \frac{2}{Z} e^{-I(t)}.\end{aligned}$$

■

We can apply Lemma 11 to prove a bound on the weighted sum of derivatives:

Lemma 12 *Let t^* be as in the definition of Theorem 5. It holds that for any $t \leq t^*$ and any sufficiently large k,*

$$\Pr\left[\left|\sum_j f'(\eta_j) u_j\right| \geq t\right] \leq e^{-t^2/2C_f k} + e^{-I(t^*)/2},$$

where C_f depends only on f.

Proof Applying Lemma 11, we have that $f'(\eta_i)$ is zero mean and is $(I(t), 2/Z_{f,1})$ bounded, where $2/Z_{f,1}$ is a constant. Since u_i and η_i are independent conditioned on $\eta_1, \ldots, \eta_{i-1}$ and since $|u_i| \leq 1$, we derive that conditioned on $\eta_1, \ldots, \eta_{i-1}$, the random variable $u_i f'(\eta_i)$ is also zero mean and $(I(t), 2/Z_{f,1})$-bounded. We would like to apply Proposition 10. Let us discuss what values we substitute in that proposition:

- First, consider the value M, that has to be lower bounded by $C\int_0^\infty (t^2+2t)e^{-I(t)/2}dt$. Notice that this integral converges due to the assumption in Theorem 5 that $I(t) \geq \Omega(\sqrt{t})$. We will substitute M to the maximum of that integral and 1
- Next, notice that we substitute n with k.
- Further, let us distinguish between the value t^* appearing in Proposition 10, that we will denote here by t', which is the solution of $t = MkI(t)/2t$, and the value appearing Theorem 5, which is the solution of $t = kI(t)/2t$, that we denote here by t^*. We note that $t' \geq t^*$, since $M \geq 1$ and since $I(t)/t$ is monotonic decreasing in t.

We obtain that for any $t \leq t^*$,

$$\Pr\left[\left|\sum_{i=1}^{k} v_i f'(\eta_i)\right| > t\right] \leq 2e^{-t^2/2Mk} + Cke^{-I(t^*)}.$$

Let us bound the second term, $Cke^{-I(t^*)}$, and for that purpose, let us obtain a lower bound on t^*: first, for a sufficiently large k, it holds that $t^* \geq 1$. Indeed, for any $t \leq 1$ and for a sufficiently large k, the value at the right hand side of the equation $t = kI(t)/2t$ is $kI(t)/2t \geq kI(1)/(2\cdot 1) \geq 1 \geq t$, which follows from the monotonicity assumption on $I(t)/t$. By definition of t^* we have $t^* \geq 1$. From the definition of t^* we have that $t^* = kI(t^*)/2t^*$, hence, $t^* = \sqrt{kI(t^*)/2}$. For any sufficiently large k, $t^* \geq 1$, hence, since $I(t)$ is increasing, we have that $t^* \geq \sqrt{kI(1)/2}$. Further, recall that $I(t) \geq \Omega(\sqrt{t})$, which implies that $I(t^*) \geq \Omega(\sqrt{k})$, hence, for a sufficiently large k,

$$Cke^{-I(t^*)} = Cke^{-I(t^*/2)} \cdot e^{-I(t^*)/2} \leq Cke^{-\Omega(\sqrt{k})} \cdot e^{-I(t^*)/2} \leq e^{-I(t^*)/2}.$$

■

Next, we would like to bound the term that corresponds to the second derivative. Recall that our goal is to bound $\sum_i |f''(\xi_i)|$ where ξ_i is in the vicinity of η_i. We start by bounding the sum $\sum_i |f''(\eta_i)|$ and then relate the sum over ξ_i to that over η_i. Since $|f''(\eta)| \leq O(f(\eta)^2)$ we can instead use the following lemma:

Lemma 13 *Let $\boldsymbol{\eta} \sim \mu_{f,1}^k$. Then, for any sufficiently large k and any $t \geq k$,*

$$\Pr\left[\left|\sum_{i=1}^{k} f(\eta_i)^2\right| > t + Ck\right] \leq e^{-c\sqrt{t}},$$

for some constants $C, c > 0$ depending only on f.

Proof We would like to apply Proposition 10, with the following substitutions:

- We replace n with k and X_i with $f(\eta_i)^2 - \mathbb{E}f(\eta_i)^2$. Notice that X_i is zero mean and $X_1, \ldots, X_k$ are independent, hence $\mathbb{E}[X_i \mid X_1 \ldots X_{i-1}] = 0$ as required.
- From Lemma 11, we have that X_i is $(\sqrt{t}, 2/Z_{f,1})$-bounded. In particular, we replace $I(t)$ with $\sqrt{t}$ and C with $2/Z_{f,1}$.
- We replace M with the corresponding integral in Proposition 10, and this integral converges as argued in Lemma 12.
- We replace t^* with with the solution of $t = Mk\sqrt{t}/2t$, and notice that $t* = (Mk/2)^{2/3}$.

Since $t^* = \Theta(k^{2/3})$ it follows that for a sufficiently large k, $k \geq t^*$. From Proposition 10 we obtain that for any $t \geq k$,

$$\Pr\left[\sum_{i=1}^{k} f(\eta_i)^2 > t + \sum_{i=1}^{k} \mathbb{E}f(\eta_i)^2\right] \leq 2e^{-c\sqrt{t}},$$

for some constant $c > 0$. Lastly, notice that from Lemma 11, we have that $\mathbb{E}[f(\eta_i)^2] < \infty$, hence, $\sum_{i=1}^{k} \mathbb{E}f(\eta_i)^2 \leq O(k)$. This concludes the proof. ■

This lets us bound $\sum_i |f(\eta_i)''|$, however, recall that we want a bound on $\sum_i |f''(\xi_i)|$ for some ξ_i in the vicinity of η_i. In fact, it suffices to bound $f(\xi_i)$ in terms of $f(\eta_i)$ and then bound $f''(\xi_i) \leq O(f(\xi_i)^2)$. Therefore, we have the following lemma:

Lemma 14 *Let $f\colon (-1,1) \to (0,\infty)$ be a function such that $\lim_{\eta\to 1^-} f(\eta) = \lim_{\eta\to -1^+} = \infty$, and $|f'(\eta)| \leq Cf(\eta)^2$ for some $C > 0$. Then, for any $\eta \in (-1,1)$ and any $\lambda \in [-1,1]$,*

$$f(\eta) \geq f\left(\eta + \frac{\lambda}{2Cf(\eta)}\right)/2.$$

Proof First, assume that $\lambda \geq 0$. Fix some η, let C be the constant such that $f'(\eta) \leq Cf(\eta)^2$ and define the function

$$g(\theta) = \frac{1}{1/f(\eta) + C(\eta - \theta)}.$$

Computing the derivative of g with respect to θ, one obtains

$$g'(\theta) = \frac{C}{(1/f(\eta) + C(\eta - \theta))^2} = Cg(\theta)^2,$$

for all θ such that $1/f(\eta) + C(\eta - \theta) > 0$. In particular, this holds for all $\theta < \eta + 1/(Cf(\eta))$. Notice that $f(\eta) = g(\eta)$ and further, that the assumption that $f'(\eta) \leq Cf(\eta)^2$ while $g'(\theta) = Cg(\theta)^2$, implies that $f(\theta) \leq g(\theta)$ for all $\theta \in [\eta, \eta + 1/(Cf(\eta)))$. In particular,

$$f\left(\eta + \frac{\lambda}{2Cf(\eta)}\right) \leq g\left(\eta + \frac{\lambda}{2Cf(\eta)}\right) = \frac{f(\eta)}{1 - \lambda/2} \leq 2f(\eta),$$

as $\lambda \leq 1$.

For the case that $\lambda < 0$, the result follows by applying the same lemma with $\tilde{\lambda} = -\lambda$, $\tilde{\eta} = -\eta$ and $\tilde{f}(\theta) = f(-\theta)$. ■

We derive the following bound on the the sum of second derivatives as a consequence:

Lemma 15 *For a sufficiently large k, and any $t \geq k$, it holds with probability at least $1-e^{-\sqrt{t}/C_f} - e^{-R/C_f}$ that*

$$\sum_{i=1}^{k} \sup_{\xi_i \colon |\xi_i - \eta_i| \leq 1/R} |f''(\xi_i)| \leq C_f t,$$

where $C_f > 0$ is a constant depending only on f.

Proof First, we bound $f''(\xi_i) \leq Cf(\xi_i)^2$, as given in the assumptions of Theorem 5. Next, we would like to bound $f(\xi_i)$ in terms of $f(\eta_i)$. Notice that $|\xi_i - \eta_i| \leq 1/R$. From Lemma 14, we have that if $1/2Cf(\eta) \geq 1/R$, then $f(\xi_i) \leq 2f(\eta_i)$. This happens whenever $f(\eta) \leq R/2C$. From Lemma 11, this happens with probability at least $1 - 2/Z_{f,1} \cdot e^{-R/2C}$. By a union bound over the k coordinates, and since $R \geq \sqrt{k}$, we have that if k is sufficiently large, then with probability $1 - e^{-R/4C}$, all the k coordinates satisfy $f(\eta) \leq R/2C$. This implies that

$$\sum_i |f''(\xi_i)| \leq C \sum_i f(\xi_i)^2 \leq 2C \sum_i f(\eta_i)^2.$$

From Lemma 13, we have that w.p. $e^{-c\sqrt{t}}$, $\sum_i f(\eta_i)^2 \leq t + O(k)$. Combining the above arguments, we obtain that with probability at least $1 - e^{-c\sqrt{t}} - e^{-R/4C}$,

$$\sum_i |f''(\xi_i)| \leq 2C \sum_i f(\eta_i)^2 \leq 2Ct + O(k) \leq O(t),$$

since we assumed in this lemma that $t \geq k$. This concludes the proof. ■

Proof [Proof of Theorem 5] From Lemma 8 it suffices to show that with probability $1 - \delta$,

$$\left| \sum_{i=1}^{k} \frac{u_i f'(\eta_i)}{R} \right| + \sum_{i=1}^{k} \max_{\xi_i \colon |\xi_i - \eta_i| \leq 1/R} \frac{|f''(\xi_i)|}{2R^2} \leq \epsilon. \tag{13}$$

The first term can be bounded by $\epsilon/2$, if we substitute $t = R\epsilon/2$ in Lemma 12, and the failure probability is bounded by

$$e^{-R^2\epsilon^2/kC_f} + e^{-I(t^*)/2}.$$

Since we assumed that $R \geq \Omega(\sqrt{k \log 1/\delta}/\epsilon)$, if the constant in the definition of R is sufficiently large, then the first term can be bounded by $\delta/4$. For the second term, recall that we assume that $\delta \geq e^{-\Omega(I(t^*))}$. If the constant in the $\Omega()$ is sufficiently small, then this term is also bounded by $\delta/4$. We conclude that the weighted sum of derivatives is bounded by $\epsilon/2$ with probability at least $1 - \delta/2$.

Next, we consider the The second term, that corresponds to the second derivatives. To bound it, we apply Lemma 15, substituting $t = C_0 k \log(1/\delta)$, where $C_0 > 0$ is a sufficiently large constant to be determined later. We derive that with probability $1 - e^{-\sqrt{t/C_f}} - e^{-R/C_f}$,

$$\sum_i |f''(\xi_i)| \leq O(k \log(1/\delta)).$$

Recall that in Eq. (13) this sum is divided by R^2. In particular, if the constant in the definition of R is sufficiently large, then this term is bounded by $\epsilon/2$ as required. Lastly, notice that the failure probability is bounded by

$$e^{-C_0\sqrt{k\log(1/\delta)}/C_f} + e^{-C_0 R/C_f} \leq 2e^{-C_0\sqrt{k\log(1/\delta)}/C_f},$$

assuming that the constant in the definition of R is sufficiently large. First, we would like to bound $k \geq \log 1/\delta$. Recall that $\delta \geq e^{-I(t^*)} \geq e^{-t^*}$, since $I(t) \leq t$. From the inequality $I(t) \leq t$ and the definition of t^*, we have that $t^* = kI(t^*)/2t^* \leq k/2 \leq k$, which, in combination with $\delta \geq e^{-t^*}$, implies that $\delta \geq e^{-k}$. Hence, $k \geq \log(1/\delta)$. Let us get back to the failure probability, and notice that it is bounded by

$$2e^{-C_0 \log(1/\delta)/C_f}.$$

Recall that C_0 is a constant that we can define, and we can set it sufficiently large such that this failure probability is at most $\delta/2$. This concludes that the bound on the sum of second derivatives is bounded by $\epsilon/2$ with probability $1-\delta/2$. In particular, Eq. (13) holds with probability $1-\delta$ which concludes the proof. ■

A.1. Proof of Proposition 10

We restate the following proposition and prove it:

Proposition 16 *Let $X_1 \cdots X_n$ be a Martingale, namely, for all i, $\mathbb{E}[X_i \mid X_1, \ldots, X_{i-1}] = 0$. Further, assume that there exists a function $I\colon [0,\infty) \to [0,\infty)$ and $C > 0$ such that*

$$\Pr[|X_i| > t \mid X_1 \cdots X_{i-1}] \leq C \exp(-I(t)),$$

and additionally, $I(t) \leq t$ and $I(t)/t$ is monotonic decreasing. Let $M > 0$ be such that $M \geq C \int_0^\infty (t^2 + 2t) e^{-I(t)/2}$ and let t^* be the unique solution to

$$t = MnI(t)/2t.$$

Then, for any $t > 0$,

$$\Pr\left[\left|\sum_{i=1}^n X_i\right| > t\right] \leq \begin{cases} 2e^{-t^2/2Mn} + Cne^{-I(t^*)} & t \leq t^* \\ 2e^{-I(t)/4} + Cne^{-I(t)} & t > t^* \end{cases}.$$

For convenience, we will refer to a random variable X as $(I(t), C)$ bounded if $\Pr[|X| > t] \leq C \exp(-I(t))$ for all $t \geq 0$.

We will use a truncation of the random variables. Given $L > 0$ define $X_i^{\leq L} = X_i \mathbb{1}(|X_i| \leq L)$ and $X_i^{>L} = X_i \mathbb{1}(|X_i| > L)$, and notice that $X_i = X_i^{\leq L} + X_i^{>L}$. We bound the sum $\sum_i X_i$ by considering the moment generating function of $\sum_i X_i^{\leq L}$ and bounding the probability that there exists i such that $X_i > L$, as formalized in the following lemma:

Lemma 17 *Fix $L, K, \theta, \delta > 0$ be such that for all i,*

$$\mathbb{E}\left[e^{\theta X_i^{\leq L}} \mid X_1 \cdots X_{i-1}\right] \leq e^K; \quad \textit{and } \Pr[|X_i| > L \mid X_1 \cdots X_{i-1}] \leq \delta.$$

Then, for any $t > 0$,

$$\Pr\left[\left|\sum_i X_i\right| > t\right] \leq 2e^{Kn - \theta t} + \delta n.$$

Proof First, by a standard induction on n, one can show that

$$\mathbb{E}\left[\exp\left(\theta\sum_{i=1}^{n} X_i^{\leq L}\right)\right] \leq e^{Kn}.$$

Consequently, by Markov's inequality,

$$\Pr\left[\sum_i X_i^{\leq L} > t\right] = \Pr\left[e^{\theta\sum_i X_i^{\leq L}} > e^{\theta t}\right] \leq \mathbb{E}[e^{\theta\sum_i X_i^{\leq L}}]/e^{\theta t} \leq e^{Kn-\theta t}.$$

Similarly, the probability that the sum is less than $-t$ can be bounded by the same quantity. Thus,

$$\Pr\left[\left|\sum_i X_i\right| > t\right] = \Pr\left[\left|\sum_i X_i^{\leq L} + X_i^{>L}\right| > t\right] \leq \Pr\left[\left|\sum_i X_i^{\leq L}\right| > t\right] + \Pr\left[\exists i,\ X_i^{>L} > 0\right]$$
$$\leq 2e^{Kn-\theta t} + n\delta.$$

■

Therefore, we would like to bound the moment generating function of X_{i+1} given $X_1 \cdots X_i$. We have the following lemma:

Lemma 18 *Let X be a zero-mean random variable. Then, for any $\theta > 0$,*

$$\mathbb{E}[e^{\theta X}] \leq 1 + \theta^2\mathbb{E}[X^2 e^{\theta|X|}]/2.$$

Proof For any $z \in \mathbb{R}$, we have by the Taylors series of e^z aronud $z = 0$,

$$e^z = 1 + z + \frac{z^2}{2}e^{\zeta(z)} \leq 1 + z + \frac{z^2}{2}e^{|z|}$$

where $\zeta(z)$ is in the segment between 0 and z. Substituting $z = \theta X$, taking expectation over X, and using the fact that X is zero mean, the result follows. ■

We would like to apply the above lemma for bounding the moment generating function.

Lemma 19 *Let X be a zero-mean random variable that is $(I(t), C)$ bounded. Then,*

$$\mathbb{E}\left[X^2 e^{\theta|X|}\right] \leq \int_0^\infty (2t + \theta t^2)e^{\theta t}\Pr[|X| > t]dt.$$

Proof Define $Z = |X|$. Then, our goal is to bound $\mathbb{E}[Z^2 e^{\theta Z}]$. By a standard change of measure argument (that can be proved, e.g., using integration by parts), for any differentiable function $h\colon [0,\infty)$ and any nonnegative r.v. Z, one has

$$\mathbb{E}[h(Z)] = h(0) + \int_0^\infty \Pr[Z > t]\frac{dh(t)}{dt}dt.$$

Applying for $h(t) = t^2 e^{\theta t}$, and substituting $dh(t)/dt = (2t + \theta t^2)e^{\theta t}$, the result follows. ■

We would like use Lemma 19 on $X^{\leq L}$:

Lemma 20 *Let X be a zero-mean random variable that is $(I(t), C)$ bounded. Let $L, \theta > 0$ such that $\theta \leq I(L)/2L$. Assume that $I(t)/t$ is monotonic decreasing and that $I(t) \leq t$. Then, for the value M defined in Proposition 10,*

$$\mathbb{E}[e^{\theta X^{\leq L}}] \leq e^{\theta^2 M/2}.$$

Proof Notice that from the monotonicity of $I(t)/t$ and from the fact that $\theta L \leq I(L)/2$, we have that for any $0 \leq t \leq L$,

$$\theta t = \theta \frac{t}{I(t)} I(t) \leq \theta \frac{L}{I(L)} I(t) \leq I(t)/2.$$

Using this inequality and the fact that $\theta \leq I(L)/2L \leq 1/2 \leq 1$, we obtain

$$\int_0^\infty (2t + \theta t^2) e^{\theta t} \Pr[|X^{\leq L}| > t] dt \leq C \int_0^L (2t + t^2) e^{\theta t - I(t)} dt \leq C \int_0^L (2t + t^2) e^{-I(t)/2} dt$$
$$\leq C \int_0^\infty (2t + t^2) e^{-I(t)/2} dt \leq M.$$

Using Lemma 18 and Lemma 19, it follows that

$$\mathbb{E}[e^{\theta X^{\leq L}}] \leq 1 + \theta^2 M/2 \leq e^{\theta^2 M/2}.$$

■

Proof [Proof of Proposition 10] We apply Lemma 17, substituting L, θ, δ, such that L is to be chosen later, $\delta = e^{-I(L)}$ and θ is a value to be chosen later that satisfies $\theta \leq \min(I(L)/2L, 1)$. From Lemma 20 we can further substitute $K = \theta^2 M/2$. We obtain that

$$\Pr\left[\left|\sum_i X_i\right| > t\right] \leq 2e^{\theta^2 Mn/2 - \theta t} + Cne^{-I(L)}. \tag{14}$$

Let us now substitute L and θ. Let t^* be the solution of $t = MnI(t)/2t$, and notice that there is a unique solution, since the left hand side (t) is increasing while the right hand side is decreasing, by the assumption that $I(t)/t$ is decreasing. If $t \leq t^*$, we take $L = t^*$ and $\theta = t/Mn$. Notice that

$$\theta = t/Mn \leq t^*/Mn = I(t^*)/2t^* = I(L)/2L,$$

as required by Lemma 20. Then,

$$e^{\theta^2 Mn/2 - \theta t} = e^{-t^2/2Mn},$$

and substituting into Eq. (14) concludes the case $t \leq t^*$. If $t > t^*$, we substitute $L = t$ and $\theta = I(t)/2t = I(L)/2L$. Then, using the definition of t^*, we can bound the right hand side of Eq. (14) by

$$2e^{\theta^* Mn/2 - \theta t} + Cne^{-I(L)} = 2e^{I(t)^2 Mn/8t^2 - I(t)/2} + Cne^{-I(t)} = 2e^{I(t)Mn/2t \cdot I(t)/4t - I(t)/2} + Cne^{-I(t)}$$
$$\leq 2e^{I(t)/4 - I(t)/2} + Cne^{-I(t)} = 2e^{-I(t)/4} + Cne^{-I(t)}.$$

This concludes the proof. ■

Appendix B. Applying the abstract bound: Proof of corollary 6

For the function $f(\eta) = 1/(1-\eta^2)^p$, one can compute that

$$\left|\frac{\partial f(\eta)}{\partial \eta}\right| = \left|\frac{2\eta p}{(1-\eta^2)^{p+1}}\right| \leq \frac{2p}{(1-\eta^2)^{p+1}} \quad .$$

Using the function $I(t) = (t/2p)^{p/(p+1)}$, we have that $I(|f'(\eta)|) \leq f(\eta)$. It is straightforward to verify that all the other conditions on $I(t)$ and f follow as well. Next, we find t^*, which is the solution to $t = kI(t)/2t$, namely to $t = k(t/2p)^{p/(p+1)}/2t$, which is solved by $t^* = C(p)k^{(p+1)/(p+2)}$, where $C(p)$ depends only on p. Lastly, we have

$$I(t^*) = C'(p) \cdot k^{p/(p+2)},$$

for some $C'(p)$, and the guarantee of the theorem implies an optimal rate for any $t \geq \exp(-I(t^*))$.

Next, we study the function $f(\eta) = \exp\left(\exp\left(1/(1-\eta^2)\right)\right)$. Denote by $h(\eta) = \frac{1}{1-\eta^2}$, and notice that $f(\eta) = \exp(\exp(h(\eta)))$. By the chain rule,

$$\begin{aligned}\left|f'(\eta)\right| = \left|\frac{d}{d\eta}\exp(\exp(h(\eta)))\right| = \left|\exp(\exp(h(\eta)))\exp(h(\eta))h'(\eta)\right| = \left|f(\eta)\log(f(\eta))\frac{2\eta}{(1-\eta^2)^2}\right| \\ \leq \left|2f(\eta)\log(f(\eta)) \cdot h(\eta)^2\right| = 2f(\eta)\log(f(\eta))(\log\log f(\eta))^2.\end{aligned}$$

Now, the intuition is to take $I(t)$ to be the inverse of $g(u) = 2u\log u \log\log^2 u$, and this will imply that $I(|f'(\eta)|) \leq f(\eta)$. This substitution, however, does not satisfy all the required assumptions on $I(t)$. To be more formal, notice that $g(u)$ is monotonic increasing in $[e,\infty)$, that $g(e) = 0$ and that $\lim_{u\to\infty} g(u) = \infty$. Hence, $g\colon [e,\infty) \to [0,\infty)$ has an inverse that we denote by $h\colon [0,\infty) \to [e,\infty)$, which is also monotonic increasing. Further, we argue that h is concave. Letting g', g'', h' and h'' denote derivatives, one has that

$$h''(t) = \frac{-1}{(g'(h^{-1}(t))^3 g''(h^{-1}(t))} < 0,$$

as g is increasing and convex in $[e,\infty)$. Since $g(u) = \omega(u)$ as $u \to \infty$, we have that $h(t) = o(t)$ as $t \to \infty$. Let $t' = \sup_t h(t) \geq t$, and define

$$I(t) = \begin{cases} t & t \leq t' \\ h(t) & t > t' \end{cases}.$$

Then, $I(t) \leq t$ as required, and further, $I(t) \geq \Omega(\sqrt{t})$ as $t \to \infty$, since $I(t) = \Theta(h(t)) = \Theta(t/(\log t \log^2\log t))$ as $t \to \infty$. It remains to argue that $I(t)/t$ is decreasing. First, $I(t)$ is a concave function as a minimum of two concave functions, and it satisfies $I(0) = 0$. Then, computing the derivative of $I(t)/t$, one has

$$\frac{d}{dt}\frac{I(t)}{t} = \frac{I'(t)t - I(t)}{t^2} = \int_0^t \frac{I'(t) - I'(s)}{t^2} ds \leq 0,$$

which follows from the fact that I is concave, hence its derivative is decreasing in t. This concludes that $I(t)/t$ is decreasing. It is straightforward to verify the other assumptions on f and $I(t)$.

Recall that $I(t) = \Theta(t/(\log t \log^2\log t))$ as $t \to \infty$. Next, we solve for t^* that is the solution of $t = kI(t)/2t = k/(2\log t \log^2\log t)$. We obtain that $t^* = \Theta(k/(\log k \log^2\log k))$, and $I(t^*) = \Theta(k/(\log^2 k \log^4\log k))$, as required.

Appendix C. Lower bound: Proof of Theorem 2

Below, we state a formal version of Theorem 2 and provide its proof.

Theorem 21 *Fix $k, \delta, \epsilon = 1$, $\Delta = 1$ and $M > 0$, and let μ be a continuous noise distribution supported on $[-M, M]$ whose density is monotonically decreasing for $\eta \geq 0$ and increasing for $\eta \leq 0$. Assume that the algorithm that adds to each answer an i.i.d. noise drawn from μ, is $(1, \delta)$ differentially private against 1-sensitive queries. Then, $M \geq \Omega(\log 1/\delta \log k)$. In particular, for $\delta = e^{-k}$, $M \geq \Omega(k \log k)$, and for any $\delta \leq e^{-\omega(k/\log^2 k)}$, $M \geq \omega(\sqrt{k \log 1/\delta})$, where $\omega()$ denotes a strict asymptotic inequality.*

We can assume that $\mu[0, \infty) \geq 1/2$ (otherwise we can consider $-\mu$ instead of μ). Further, recall that μ is a continuous distribution, hence it has density that we can denote by $e^{-f(x)}$, where $f(x) = \infty$ if the density is zero. Our assumption implies that $f(\eta)$ is increasing for $\eta \geq 0$ and decreasing for $\eta \leq 0$.

The argument consists of the following lemmas:

Lemma 22 *Assume that $\Pr_\mu[\eta \geq 0] \geq 1/2$ and that $\delta \leq 0.1$. Then,*

$$f(1/2) \leq \log(10M).$$

For an intuitive explanation about this lemma, notice that it is stating that the density at $1/2$ cannot be very small. This follows from two facts: (1) the density at 0 is at least $1/2M$, since $\mu([0, M]) \geq 1/2$ and the density is decreasing in $[0, M]$. Further, since the noise satisfies $(1, 0.1)$-DP, its density cannot drop "too fast", other a change in the true query value will be detected. The formal proof of this lemma appears in Section C.1.

Lemma 23 *Let $\eta \geq 1/2$ such that*

$$\max\left(\log 2, \frac{\log(1/2\delta)}{4(k-1)}\right) \leq f(\eta) \leq \frac{\log(1/2\delta)}{3}.$$

Then,

$$f(\eta + 1/2) \leq f(\eta) \exp(8/\log(2/\delta)).$$

To gain some intuition on this lemma, we again use the fact that since μ satisfies DP, the noise density cannot drop too fast. In particular, if the density at η is non-negligible, then the density at $\eta + 1/2$ cannot drop too fast. Notice that the assumption that the density at η is non-negiligible corresponds to requiring that $f(\eta) \leq \log(1/\delta)/3$. On the other hand, the lower bound requirement on f weakens as k grows. This is due to the fact that we utilize the multiple samples. The proof appears in Section C.2.

The proof concludes by combining these two lemmas. Since $f(1/2)$ is relatively and due to the bound on the growth rate of f, we conclude that $f(\eta) \leq O(\log(1/\delta))$ for some $\eta \geq \Omega(\log k \log 1/\delta)$. In particular, this implies that $[0, \eta]$ is contained in the support of μ, and concludes the proof. The analysis is based on a case analysis as formalized below:

Proof [Proof of Theorem 21] To complete the proof, let η_0 be the minimal η such that $\eta \geq 1/2$ and

$$f(\eta) \geq \max\left(\log 2, \frac{\log(2/\delta)}{4(k-1)}\right).$$

Then, by Lemma 22,

$$f(\eta_0) \leq \max\left(\log 2, \frac{\log(2/\delta)}{4(k-1)}, \log(10M)\right). \tag{15}$$

Applying Lemma 23 multiple times, we derive that

$$f(\eta_0 + i/2) \leq f(\eta_0)e^{i\cdot 8/\log(2/\delta)},$$

for any i such that

$$f(\eta_0)e^{i\cdot 8/\log(2/\delta)} \leq \frac{\log(2/\delta)}{3}.$$

Equivalently, this holds for any

$$i \leq \ell := \frac{\log(\log(2/\delta)/3f(\eta_0))}{8/\log(2/\delta)} = \log\left(\frac{\log(2/\delta)}{3f(\eta_0)}\right)\cdot\frac{\log(2/\delta)}{8} \quad .$$

In particular, $f(\eta_0 + \lfloor\ell\rfloor/2) < \infty$, which implies that $M > \eta_0 + \lfloor\ell\rfloor/2$. It suffices to show that $\ell \geq \Omega(\log 1/\delta\cdot\log k)$ to conclude the proof. We divide into cases according to $f(\eta_0)$, using Eq. (15).

- If $f(\eta_0) \leq \log 2$: then, $\ell \geq \Omega(\log(2/\delta)\log\log(2/\delta))$. We divide into cases according to δ: if $\delta \leq e^{-k/\log^2 k}$ then $\log\log(2/\delta) \geq \Omega(\log k)$ and the proof follows. Otherwise, we use the theorem of Steinke and Ullman (2016) that claims that for any $\delta \geq e^{-k}$ and for any $(1,\delta)$ mechanism for 1-sensitive queries, it holds that the average error is bounded as follows:

  $$\frac{1}{k}\sum_{i=1}^{k}|q_i(\mathbf{x}) - a_i| \geq \Omega(\sqrt{k\log(1/\delta)}) \quad .$$

  This implies that if the mechanism uses independent bounded noise of magnitude bounded by M, then $M \geq \Omega(\sqrt{k\log(1/\delta)})$. Since $\sqrt{k\log(1/\delta)} \geq \log 1/\delta\log k$ for $\delta \leq e^{-k/\log^2 k}$, the result follows.

- If $f(\eta_0) \leq \log(10M)$. As argued for the previous case, we can assume that $\delta \leq e^{-k/\log^2 k}$. Further, we can assume that $M \leq \log(1/\delta)\log(k)/10$, otherwise the theorem follows. The above two assumptions imply that

  $$f(\eta_0) \leq \log(10M) \leq \log\log(1/\delta) + \log\log k \leq O(\log\log(1/\delta)).$$

  This implies that

  $$\ell = \log\left(\frac{\log(2/\delta)}{3f(\eta_0)}\right)\cdot\frac{\log(2/\delta)}{8} \geq \Omega\left(\log\log(1/\delta)\log(1/\delta)\right) \geq \Omega(\log k\log(1/\delta)),$$

  where the last inequality follows from $\delta \leq e^{-k/\log^2 k}$.

- If $f(\eta_0) \leq \log(2/\delta)/4(k-1)$, it clearly follows from definition of ℓ that $\ell \geq \Omega(\log(1/\delta)\log k)$.

This concludes the proof. ■

C.1. Proof of Lemma 22

We start with a simple lemma that argues that if we shift any subset U of $\mathbb{R}^k$ by any $v \in [0, \Delta]^k$ then the probability of the shifted set should not significantly differ from that of U, if the noise satisfies DP.

Lemma 24 *Let ϵ, δ, let μ^k be a noise that is (ϵ, δ) differentially private against 1-sensitive queries. Let $U \subseteq \mathbb{R}^k$, $v \in [0, 1]^k$, and define $U + v = \{u + v \colon u \in U\}$. Then.*

$$\mu^k(U) \leq e^{\epsilon} \mu^k(U + v) + \delta.$$

Proof Assume that $\mathcal{X} = [-1, 0]$, that $n = 1$, and define the query $q_i(x) = x$ for all $i = 1, \ldots, k$. Let $x_1 = (0, \ldots, 0)$ and $x_1' = -v$. Recall that $a_1, \ldots, a_k$ are the answers of the algorithm, and notice that by the (ϵ, δ) differential privacy,

$$\Pr[(a_1, \ldots, a_k) \in U \mid x_1] \leq e^{\epsilon} \Pr[(a_1, \ldots, a_k) \in U \mid x_1'] + \delta.$$

This is equivalent to

$$\Pr_{\mu}[\eta \in U] \leq e^{\epsilon} \Pr_{\mu}[\eta \in U + v] + \delta,$$

as required. ■

We can conclude with the proof.

Proof [Proof of Lemma 22] Assume towards contradiction that $f(1/2) > \log(10M)$, and this implies by monotonicity that $f(1) > \log(10M)$. Then,

$$\mu[1, \infty) = \mu[1, M] = \int_1^M e^{-f(\eta)} d\eta \leq \int_1^M e^{-f(1)} d\eta \leq M e^{-f(1)} < 1/10.$$

Consequently,

$$\mu[0, 1] = \mu[0, \infty) - \mu[1, \infty) > 0.4$$

while

$$\mu[1, 2] \leq \mu[1, \infty) < 0.1.$$

Let $U = \{(x_1, \ldots, x_n) \colon x_1 \leq 1\}$. It holds that

$$\mu^n(U) = \mu[0, 1] > 0.4,$$

while

$$\mu^n(U + (1, 0, \ldots, 0)) = \mu[1, 2] < 0.1.$$

This implies that

$$\mu^n(U) > e^1 \mu^n(U + (1, 0, \ldots, 0)) + 0.1,$$

which contradicts Lemma 24 and the fact that the noise is $(1, 0.01)$ private. ■

C.2. Proof of Lemma 23

First we use the following result, which is analogous to a bound that appears in the upper bound in this paper, and follows from Lemma 24 above.

Lemma 25 *If the noise if μ^k is $(1,\delta)$ private with respect to 1-sensitive queries, then for any $v_1,\ldots,v_k \in [0,1]$,*

$$\Pr_{x\sim\mu}\left[\sum_{i=1}^{k} f(x_i+v_i)-f(x_i)\geq 2\right] < 2\delta. \tag{16}$$

Proof Look at the set $U=\{\eta\colon \sum_i f(\eta_i+v_i)-f(\eta_i)\geq 2\}$. From Lemma 24 and the $(1,\delta)$ privacy assumption, we have

$$\Pr_{\mu^k}[U]\leq e\Pr_{\mu^k}[U+v]+\delta.$$

Further,

$$\Pr_{\mu^k}[U]=\int_U \exp\left(-\sum_{i=1}^{k} f(\eta_i)d\eta\right)\geq \int_U \exp\left(-2-\sum_{i=1}^{k} f(\eta_i+v_i)d\eta\right)=e^2\Pr_{\eta}[U+v].$$

Combining the above inequalities, we derive that

$$\Pr_{\mu}[U]\leq e\Pr_{\mu}[U+v]+\delta\leq \Pr_{\mu}[U]/e+\delta,$$

hence

$$\Pr_{\eta}[U](1-1/e)\leq\delta,$$

which implies that $\Pr_\eta[U]<2\delta$ as required. ■

Recall that we want to bound $f(\eta+1/2)-f(\eta)$. In the proof, we assume towards contradiction that $f(\eta_0+1/2)-f(\eta_0)$ is large for some appropriate η_0 and we will derive that Eq. (16) fails to hold, which, by Lemma 25 implies that the mechanism is not DP. As a first step, we will prove that if $f(\eta_0+1/2)-f(\eta_0)$ is large then a variant of Eq. (16) is not satisfied, with different constants and $k=1$.

Claim 26 *Let $\eta_0\geq 1/2$. Then,*

$$\Pr_{\eta\sim\mu}\left[f(\eta+1)-f(\eta)\geq f(\eta_0+1/2)-f(\eta_0)\right]\geq\frac{e^{-f(\eta_0)}}{2}\ .$$

Proof We have by monotonicity of $f(\eta)$,

$$\Pr_{\eta\sim\mu}[\eta_0-1/2\leq\eta\leq\eta_0]=\int_{\eta_0-1/2}^{\eta_0} e^{-f(\eta)}d\eta\geq\int_{\eta_0-1/2}^{\eta_0} e^{-f(\eta_0)}d\eta=\frac{e^{-f(\eta_0)}}{2}\ .$$

For any such $\eta\in[\eta_0-1/2,\eta_0]$ we have

$$f(\eta)\leq f(\eta_0)\leq f(\eta_0+1/2)\leq f(\eta+1).$$

Consequently, $f(\eta+1)-f(\eta) \geq f(\eta_0+1/2)-f(\eta_0)$. ■

Next, we extend Claim 26 to show that if $f(\eta_0+1/2)-f(\eta_0)$ is large for some η_0, then a variant of Eq. (16) does not hold, where the sum is over $m>1$ elements. To achieve that, we first use Claim 26 to show that $\Pr[f(\eta+1)-f(\eta) \geq a] \geq b$ for some $a, b>0$ and then derive that if $\eta_1, \ldots, \eta_m$ are i.i.d., then

$$\Pr[\sum_{i=1}^{m} f(\eta_i+1)-f(\eta_i) \geq ma] \geq \Pr[\forall i \leq m, f(\eta_i+1)-f(\eta_i) \geq a]$$
$$= \prod_{i=1}^{m} \Pr[f(\eta_i+1)-f(\eta_i) \geq a] = b^m.$$

Choosing a and b appropriately yields the desired result.

Lemma 27 *Let $\eta_0 \geq 1/2$, let $\delta_0>0$, and let $C>0$. Assume that*

$$f(\eta_0+1/2) \geq f(\eta_0)(1+4C/\log(1/\delta_0))$$

and that

$$\max\left(\log 2, \frac{\log(1/\delta_0)}{4(k-1)}\right) \leq f(\eta_0) \leq \frac{\log(1/\delta_0)}{3}.$$

Then, there is $m \leq k$ such that

$$\Pr_{\eta \sim \mu^k}\left[\sum_{i=1}^{m} f(\eta_i+1)-f(\eta_i) \geq C\right] \geq \delta_0.$$

Proof Define $K=4Cf(\eta_0)/\log(1/\delta_0)$ and $L=f(\eta_0)+\log 2$. Applying Claim 26, we have

$$\Pr_{\eta \sim \mu}[f(\eta+1)-f(\eta) \geq K] \geq \Pr_{\eta \sim \mu}[f(\eta+1)-f(\eta) \geq f(\eta_0+1/2)-f(\eta_0)] \geq e^{-L}.$$

Let $m=\lceil C/K \rceil$. First, notice that $m \leq k$: indeed, it suffices to show that $C/K+1 \leq k$, which holds since

$$f(\eta_0) \geq \frac{\log 1/\delta_0}{4(k-1)} \quad .$$

Then,

$$\Pr_{\eta \sim \mu^k}[\sum_{i=1}^{m} f(\eta_i+1)-f(\eta_i) \geq C] \geq \Pr_{\eta \sim \mu^k}[\forall i \in \{1, \ldots, m\},\ f(\eta_i+1)-f(\eta_i) \geq K] \geq e^{-Lm}.$$

It remains to argue that $e^{-Lm} \geq \delta_0$, or equivalently, $Lm \leq \log(1/\delta_0)$. By definition of m, it suffices to show that $L(C/K+1) \leq \log 1/\delta_0$. Indeed, using the fact that $f(\eta_0) \geq \log 2 \geq 1/2$ and that $f(\eta_0) \leq \log(1/\delta_0)/3$,

$$L(C/K+1) = LC/K+L = \log 1/\delta_0 \frac{f(\eta_0)+\log 2}{4f(\eta_0)} + f(\eta_0)+\log 2$$
$$\leq \log 1/\delta_0 \frac{2f(\eta_0)}{4f(\eta_0)} + 2f(\eta_0) \leq 3f(\eta_0) \leq \log(1/\delta_0)$$

This concludes the proof. ■

Proof [Proof of Lemma 23] Assume towards contradiction the existence of such η. Then, since $e^x \geq 1 + x$ for all x, we have

$$f(\eta + 1/2) \geq f(\eta)(1 + 8/\log(2/\delta)).$$

Applying Lemma 27 with $C = 2$ and $\delta_0 = 2\delta$, we obtain that there eixsists $m \leq k$ such that

$$\Pr_{\eta \sim \mu^k}\left[\sum_{i=1}^{m} f(\eta_i + 1) - f(\eta_i) \geq 2\right] \geq 2\delta.$$

However, by Lemma 25 this does not hold. We derive the contradiction, and this concludes the proof. ■

Appendix D. Adaptive data analysis: Proof of Corollary 3

We use the following transfer theorem from Jung et al. (2021):

Theorem 28 *Assume that $\mathcal{A}$ is an algorithm that answers k statistical queries, $q_i \colon X \to [0,1]$ given some dataset $(x_1, \ldots, x_n) \in X^n$. Further, assume that the algorithm is (ϵ, δ)-differentially private with respect to its dataset and that with probability $1 - \beta'$, all of its answers a_i are α'-accurate with respect to the sample, namely,*

$$\Pr\left[\forall i = 1, \ldots, k \colon \left| a_i - \frac{1}{n}\sum_{j=1}^{n} q_i(x_j) \right| \leq \alpha'\right] \geq 1 - \beta' \ .$$

Assume that $x_1, \ldots, x_n$ are drawn from some distribution P. Then, for any $c, d > 0$, the algorithm M produces answers that are $\alpha(c,d)$-accurate with respect to P with probability $1 - \beta(c,d)$, where

$$\alpha(c,d) = \alpha' + e^{\epsilon} - 1 + c + 2d; \quad \beta(c,d) = \beta'/c + \delta/d.$$

Namely,

$$\Pr\left[\forall i = 1, \ldots, k \colon |a_i - q_i(P)| \leq \alpha(c,d)\right] \geq 1 - \beta(c,d) \ .$$

From this, we can easily derive our theorem:

Proof [Proof of Corollary 3] Fix $\alpha, \beta \in (0, 1/2)$. We apply Theorem 28, using the bounded noise mechanism from Theorem 1, that answers each query i with $a_i = \sum_{i=j}^{n} q_i(x_j) + \eta_i$ where η_i are i.i.d. bounded noise. We set the privacy parameters to $\epsilon = \alpha/8$ and $\delta = \alpha\beta/4$. We obtain that the answers are α'-accurate with respect to the sample, for

$$\alpha' = O\left(\frac{\sqrt{k\log(1/\alpha\beta)}}{\epsilon n}\right),$$

using Theorem 1 and the fact that the statistical queries on a dataset of size n are $\Delta = 1/n$-sensitive. This holds with probability 1, hence, we can substitute $\beta' = 0$. We take $c \to 0$ and $d = \alpha/4$, and we derive that

$$\lim_{c \to 0} \alpha(c,d) = \alpha' + e^{\alpha/8} - 1 + \alpha/2 \leq \alpha' + 3\alpha/4; \quad \forall c > 0,\ \beta(c,d) = \beta.$$

Here, we used $e^x \leq 1 + 2x$ for $x \in [0, 1]$. If we take

$$n = \Theta\left(\frac{\sqrt{k\log(1/\delta)}}{\epsilon^2}\right) = \Theta\left(\frac{\sqrt{k\log(1/\alpha\beta)}}{\alpha^2}\right),$$

then we have $\alpha' < \alpha/4$, hence, $\alpha(c, d) < \alpha$ for some $c > 0$ and $\beta(c, d) = \beta$. We derive by Theorem 28 that the protocol is α-accurate with respect to P with probability $1 - \beta$, as required. ■

Appendix E. Computing tighter upper bounds

Here we explain how to derive an algorithm that upper bounds the optimal noise level, for each given ϵ, k and δ. The final algorithm is given as Algorithm 2 below, yet, we start by explaining it step by step. First of all, we note the following sufficient condition for (ϵ, δ)-privacy, which is a tighter variant of its analogoue in the proof of Theorem 5:

Lemma 29 *Let P and Q be probability distributions over $\mathbb{R}^k$ with densities $p(x)$ and $q(x)$, respectively. Let $\epsilon, \delta > 0$, and assume that*

$$\int_{\epsilon}^{\infty} \Pr_{X\sim P}\left[\log\frac{p(X)}{q(X)} > t\right] e^{\epsilon - t} dt \leq \delta.$$

Then, for any $U \subseteq \mathbb{R}^d$,

$$\Pr_{X\sim P}[X \in U] \leq e^{\epsilon} \Pr_{X\sim Q}[X \in U] + \delta.$$

Proof First of all, notice that by change of variables $s = t - \epsilon$, one has

$$\int_{0}^{\infty} \Pr_{X\sim P}\left[\log\frac{p(X)}{q(X)} - \epsilon > s\right] e^{-s} ds \leq \delta \ .$$

Assume that the above statement holds and denote by Λ the random variable

$$\Lambda = \max\left(0, \log\frac{p(X)}{q(X)} - \epsilon\right)$$

where $X \sim P$. The left hand side translates to

$$\int_{0}^{\infty} \Pr[\Lambda > t] e^{-t} dt \leq \delta.$$

We use the known technique of integration by parts for probability distributions, which states that for a nonnegative random variable Z and for a function $F\colon [0, \infty) \to \mathbb{R}$ with a continuous derivative that satisfies $F(0) = 0$,

$$\mathbb{E}[F(Z)] = \int_{0}^{\infty} \Pr[Z > t] F'(t) dt.$$

Substituting $Z = \Lambda$ and $F(t) = 1 - e^{-t}$, we derive that

$$\mathbb{E}[1 - e^{-\Lambda}] \leq \delta.$$

Using the fact that $1 - e^{-0} = 0$ and the fact that $\Lambda \geq 0$, we derive that

$$\mathbb{E}[(1 - e^{-\Lambda})\mathbb{1}(\Lambda > 0)] \leq \delta.$$

Substituting the value of Λ, we obtain

$$\int_{\mathbb{R}^k} p(x) \left(1 - e^{\epsilon} \frac{q(x)}{p(x)}\right) \mathbb{1}\left(\log \frac{p(x)}{q(x)} > \epsilon\right) dx = \int_{\mathbb{R}^k} (p(x) - e^{\epsilon} q(x))\, \mathbb{1}\left(p(x) > e^{\epsilon} q(x)\right) dx \leq \delta.$$

This implies that for any $U \subseteq \mathbb{R}^k$,

$$\begin{aligned}\int_U p(x) - e^{\epsilon} q(x) dx &\leq \int_U (p(x) - e^{\epsilon} q(x))\, \mathbb{1}\left(p(x) > e^{\epsilon} q(x)\right) dx \\ &\leq \int_{\mathbb{R}^k} (p(x) - e^{\epsilon} q(x))\, \mathbb{1}\left(p(x) > e^{\epsilon} q(x)\right) dx \leq \delta,\end{aligned}$$

as required. ■

Let us apply the above lemma for answering multiple queries. Below, we assume for simplicity that the queries are non-interactive, namely, $q_1, \ldots, q_k$ are given a-priori. Yet, one can obtain the *exact* same bounds while assuming that they are asked adaptively. We refer to the proof of Theorem 5 (and particularly, to Lemma 17).

Lemma 30 *Let M be a mechanism that answers k fixed Δ-sensitive queries by adding an i.i.d. noise drawn from some distribution μ over $\mathbb{R}$, with density*

$$\mu(\eta) = \frac{e^{-f(\eta)}}{Z}, \quad \text{where } Z = \int_{\mathbb{R}} e^{-f(\eta)} d\eta.$$

(Here, we use the convention $f(\eta) = \infty$ if $\mu(\eta) = 0$.) Let $\epsilon, \delta > 0$ and assume that for all $v_1, \ldots, v_k \in [-\Delta, \Delta]$, it holds that

$$\int_{\epsilon}^{\infty} \Pr_{\eta_1, \ldots, \eta_k \overset{\text{i.i.d.}}{\sim} \mu} \left[\sum_{i=1}^{k} f(\eta_i + v_i) - f(\eta_i) > t\right] e^{\epsilon - t} dt \leq \delta. \tag{17}$$

Then, the mechanism is (ϵ, δ)-private.

Proof Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be two neighboring datasets, let $q_1, \ldots, q_k$ denote the queries. Denote the densities of the output of the mechanism on $\mathbf{x}_1$ and $\mathbf{x}_2$ by p_1 and p_2, respectively. Our goal is to show that for all $U \subseteq \mathbb{R}^d$,

$$\int_U p_1(y_1, \ldots, y_k) dy \leq e^{\epsilon} \int_U p_2(y_1, \ldots, y_k) dy + \delta.$$

In order to show this inequality, from Lemma 29 it suffices to show that

$$\int_{\epsilon}^{\infty} \Pr_{(y_1, \ldots, y_k) \sim p_1} \left[\log \frac{p_1(y_1, \ldots, y_k)}{p_2(y_1, \ldots, y_k)} > t\right] e^{\epsilon - t} dt \leq \delta. \tag{18}$$

Towards this goal, notice that

$$p_j(y_1, \ldots, y_k) = \prod_{i=1}^{k} \mu(y_i - q_i(\mathbf{x}_j)).$$

In particular,

$$\log \frac{p_1(y_1, \ldots, y_k)}{p_2(y_1, \ldots, y_k)} = \sum_{i=1}^{k} f(y_i - q_i(\mathbf{x}_2)) - f(y_i - q_i(\mathbf{x}_1)). \tag{19}$$

Assuming that $(Y_1, \ldots, Y_k)$ is a random variable denoting the output of the mechanism on $\mathbf{x}_1$, we have that $Y_i - q_i(\mathbf{x}_1)$ are distributed i.i.d. according to μ. Hence, the right hand side of Eq. (19) is distributed according to

$$\sum_{i=1}^{k} f(\eta_i + q_i(\mathbf{x}_1) - q_i(\mathbf{x}_2)) - f(\eta_i),$$

where $\eta_1, \ldots, \eta_k \overset{\text{i.i.d.}}{\sim} \mu$. Denote $v_i = q_i(\mathbf{x}_1) - q_i(\mathbf{x}_2)$ and since the queries are Δ-sensitive, $|v_i| \leq \Delta$. We have that the right hand side of Eq. (19) equals

$$\sum_{i=1}^{k} f(\eta_i + v_i) - f(\eta_i).$$

Combining with the assumption of this lemma, this proves Eq. (18), which concludes the proof. ■

Next, we show how to bound the deviations of $\sum_i f(\eta_i + v_i) - f(\eta_i)$. A standard way is to use the moment generating function, however, for the bounded noises suggested in this paper, the corresponding MGF might not exist. Instead, one can use truncation. In particular, we will find some threshold L such that $\Pr[|\eta_i| > L] \leq \delta_1/k$, for some $\delta_1 < \delta$. This implies that $\Pr[\exists i = 1, \ldots, k\colon\ |\eta_i| > L] \leq \delta_1$. Then, the left hand side of Eq. (17) can be bounded by

$$\int_{\epsilon}^{\infty} \left(\Pr_{\eta_1, \ldots, \eta_k \overset{\text{i.i.d.}}{\sim} \mu} \left[\sum_{i=1}^{k} f(\eta_i + v_i) - f(\eta_i) > t \ \wedge \ \max_i |\eta_i| \leq L \right] + \Pr[\exists i,\ |\eta_i| > \ell] \right) e^{\epsilon - t} dt$$
$$\leq \int_{\epsilon}^{\infty} \Pr_{\eta_1, \ldots, \eta_k \overset{\text{i.i.d.}}{\sim} \mu} \left[\sum_{i=1}^{k} f(\eta_i + v_i) - f(\eta_i) > t \ \wedge \ \max_i |\eta_i| \leq L \right] e^{\epsilon - t} dt + \delta_1,$$

using the fact that $\Pr[\exists i,\ |\eta_i| > \ell] \leq \delta_1$ and $\int_{\epsilon}^{\infty} e^{\epsilon - t} = 1$. Denote $X_i = (f(\eta_i + v_i) - f(\eta_i))\mathbb{1}(|\eta_i| \leq L)$ and denote $\delta_2 = \delta - \delta_1$. It is sufficient to prove that

$$\int_{\epsilon}^{\infty} \Pr\left[\sum_i X_i > t \right] e^{\epsilon - t} dt \leq \delta_2. \tag{20}$$

We bound the deviations of $\sum_i X_i$ using the moment generating function technique: for any $\lambda > 0$, by Markov's inequality, one has

$$\Pr\left[\sum_i X_i > t \right] = \Pr\left[e^{\lambda \sum_i X_i} > e^{\lambda t} \right] \leq \frac{\mathbb{E}\left[e^{\lambda \sum_i X_i} \right]}{e^{\lambda t}} = \frac{\prod_i \mathbb{E}\left[e^{\lambda X_i} \right]}{e^{\lambda t}}. \tag{21}$$

Recall that $X_i = (f(\eta_i + v_i) - f(\eta_i))\mathbb{1}(|\eta_i| \leq L)$. We would like to eliminate the dependence on $v_i \in [-\Delta, \Delta]$. We will bound its MGF as follows:

Lemma 31 *Assume that μ is distribution with density $\mu(\eta) \propto e^{-f(\eta)}$ defined on $[-R, R]$. Assume that μ is log concave, namely, $\mu((\eta_1+\eta_2)/2) \geq \sqrt{\mu(\eta_1)\mu(\eta_2)}$. Further assume that μ is symmetric, namely, $\mu(\eta) = \mu(-\eta)$. Let $\lambda, L > 0$. Define for any $|t| < R - L$ the random variable*

$$X_t = (f(\eta + t) - f(\eta))\mathbb{1}(|\eta| \leq L),$$

where $\eta \sim \mu$. Then, for any $|a| \leq |b| < R - L$,

$$\mathbb{E}[\exp(\lambda X_a)] \leq \mathbb{E}[\exp(\lambda X_b)]$$

(notice that X_t is undefined for $|t| > R - L$).

Proof It is sufficient to assume that f has a continuous derivative. Otherwise, one can approximate f with a sequence of functions with continuous derivatives. We will show that $\mathbb{E}[e^{\lambda X_t}]$ is monotone increasing when $t \geq 0$, and the result will follow since $\mathbb{E}[e^{\lambda X_t}] = \mathbb{E}[e^{\lambda X_{-t}}]$, as μ is symmetric. Let E be the event that $|\eta| \leq L$. Then, from symmetricity of μ and f,

$$\mathbb{E}\left[e^{\lambda X_t}\right] = \mathbb{E}\left[e^{\lambda f(\eta+t)}e^{-\lambda f(\eta)}\mathbb{1}_E\right] = \mathbb{E}\left[\frac{e^{\lambda f(\eta+t)} + e^{\lambda f(-\eta+t)}}{2}e^{-\lambda f(\eta)}\mathbb{1}_E\right]$$
$$\mathbb{E}\left[\frac{e^{\lambda f(|\eta|+t)} + e^{\lambda f(|\eta|-t)}}{2}e^{-\lambda f(\eta)}\mathbb{1}_E\right].$$

It is sufficient to show that the following derivative is nonnegative, for any $\eta \geq 0$:

$$\frac{d}{dt}\frac{e^{\lambda f(\eta+t)} + e^{\lambda f(\eta-t)}}{2} = \frac{f'(\eta+t)e^{\lambda f(\eta+t)} - f'(\eta-t)e^{\lambda f(\eta-t)}}{2}. \tag{22}$$

Since μ is symmetric and log-concave, then f is convex and symmetric. In particular, this implies that f has a minimum at 0 and it is monotonic nondecreasing at $\eta > 0$. Since we assumed that $\eta, t \geq 0$, this implies that $\eta + t \geq |\eta - t|$, which implies that $f(\eta+t) \geq f(\eta-t)$. Further, convexity of f implies that its derivative is increasing, which implies that

$$f'(\eta + t) \geq f'(|\eta - t|) = |f'(\eta - t)|,$$

using the fact that the derivative of a symmetric function is antisymmetric. The above implies that

$$f'(\eta + t)e^{\lambda f(\eta+t)} \geq \left|f'(\eta - t)e^{\lambda f(\eta-t)}\right|,$$

which derives that Eq. (22) is nonnegative and concludes the proof. ■

Define

$$X := (f(\eta + \Delta) - f(\eta))\mathbb{1}(|\eta| \leq L),$$

where $\eta \sim \mu$, and the above lemma implies that $\mathbb{E}[e^{\lambda X_i}] \leq \mathbb{E}[e^{\lambda X}]$ for all $i = 1, \ldots, k$. Combining with Eq. (21), one has

$$\Pr\left[\sum_i X_i > t\right] \leq \inf_{\lambda>0} \frac{\mathbb{E}\left[e^{\lambda X}\right]^k}{e^{\lambda t}} \leq \exp\left(\inf_{\lambda>0} k \log\left(\mathbb{E}\left[e^{\lambda X}\right]\right) - \lambda t\right).$$

The function $k \log \left(\mathbb{E}\left[e^{\lambda X}\right]\right) - \lambda t$ is known to be convex in λ for any random variable X, hence, it can be optimized efficiently, and any λ would yield an upper bound. Integrating over t, one can bound the left hand side of Eq. (20). This produces a way to certify that Eq. (20) holds, for given ϵ, δ, k and μ. Recall that if this inequality holds, then the mechanism is guaranteed to be (ϵ, δ)-DP. In particular, given f and R, Eq. (20) certifies that $M_{f,R}$ is (ϵ, δ)-DP. If we want to find an upper bound on the minimal magnitude R such that $M_{f,R}$ is (ϵ, δ)-DP, we can perform a simple binary search over values of $R > 0$ (stopping when the desired precision has achieved). We note that in order to obtain a proper upper bound, one has to ensure that the approximation errors in the relevant computations are one sided (e.g. the computed MGF value should not be lower than the actual value). Algorithm 1 tests if a mechanism is (ϵ, δ)-DP and Algorithm 2 finds an upper bound on the minimal noise R given some function f.

Function `testPrivacy`$(\epsilon, \delta, k, \Delta, p)$**:** `/* Checks if a mechanism is` (ϵ, δ)`-private */`

Input: Privacy parameters $\epsilon > 0$ and $\delta \in (0, 1)$; Number of queries $k \in \mathbb{N}$; sensitivity $\Delta > 0$; A probability density p over $(-R, R)$ such that $\log p(y)$ is concave and $p(y) = p(-y)$.

Output: An indication whether the mechanisms that answers k Δ-sensitive queries with i.i.d. noise according to p is (ϵ, δ)-DP. A **false** answer may be wrong but a true answer is always correct.

Function `MGF`(L, λ)**:** `/* Computes a moment generating function */`

Input: A threshold $L > 0$ and a real parameter $\lambda > 0$

Output: The moment generating function $\mathbb{E}[e^{\lambda X}]$, where $Y \sim p$ and $X = (\log p(Y) - \log p(Y + \Delta))\mathbb{1}(|Y| \leq L)$

return $\int_{-L}^{L} p(y) \exp(\lambda(\log p(y) - \log p(y + \Delta)))dy$ `/* An upper bound is also valid and can be computed since` $\log p(y) - \log p(y + \Delta)$ `is monotone increasing. */`

return

Function `deviationBound`(L, t)**:** `/* Computes a probability of deviation */`

Input: Real numbers $L, t > 0$

Output: An upper bound on the probability that $\sum_{i=1}^{k} X_i > t$, where $Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} p$, $v_i \in [-\Delta, \Delta]$ are arbitrary and $X_i = (\log p(Y_i) - \log p(Y_i + v_i))\mathbb{1}(|Y_i| \leq L)$

logProb $\leftarrow \inf_{\lambda > 0} k\texttt{MGF}(L, \lambda) - \lambda t$ **return** e^{logProb}

return

Tunable Parameter: $\delta_1 \in (0, \delta)$.

`/* Can be set arbitrarily. A possible setting is:` $\delta_1 = 0.01\delta$ `*/`

$L \leftarrow$ the unique value in $[0, R]$ such that $\int_{-L}^{L} p(y)dy = 1 - \delta_1$ `/* An upper bound is also valid. */` **if** $L + \Delta \geq R$ **then**

return *false*

end

$\delta_2 \leftarrow \int_{\epsilon}^{\infty}$ `deviationBound`$(L, t)e^{\epsilon - t}dt$ `/* An upper bound is valid and can be computed since deviationBound is monotone decreasing in` t `*/`

if $\delta_1 + \delta_2 \leq \delta$ **then**

return *true*

else

return *false*

end

Algorithm 1: Check if a mechanism satisfies DP

return

Function `noiseUpperBound(`$\epsilon, \delta, k, \Delta, p_1$`):` `/* Computes an upper bound on the noise required for preserving a desired privacy level */`

Input: Privacy parameters $\epsilon > 0$ and $\delta \in (0,1)$; Number of queries k; Sensitivity Δ; A probability density function $p_1 \colon (-1,1) \to (0,\infty)$ such that $\log p_1$ is concave and $p_1(y) = p_1(-y)$.

Definition: For all $R > 0$, define by p_R the density over $[-R, R]$, such that $p_R(y) = p_1(y/R)/R$.

`/* Equivalently, a sample` $y \sim p_R$ `is obtained by sampling` $y' \sim p_1$ `and outputting` Ry'`. */`

Output: A number $R > 0$ such that p_R is (ϵ, δ)-DP for answering k Δ-sensitive queries.

err$\leftarrow$ a very small number `/* For example,` 10^{-8} `*/`

`/* Compute an upper bound` b `on the minimal allowable noise` R`. */`

$b \leftarrow 1$ **while not** `testPrivacy(`$\epsilon, \delta, k, \Delta, p_R$`)` **do**

  $b \leftarrow 2 * b$

end

$a \leftarrow 0$ `/* A lower bound on the minimal allowable noise` R `*/`

while $b - a > err$ **do**

  $m \leftarrow (a+b)/2$ **if** `testPrivacy(`$\epsilon, \delta, k, \Delta, p_m$`)` **then**

    $b \leftarrow m$

  else

    $a \leftarrow m$

  end

end

return b

return

Algorithm 2: Find a suitble noise magnitude R