Generalized Private Selection and Testing with High Confidence

Edith Cohen
Google Research, Mountain View, CA, USA
Tel Aviv University, Israel

Xin Lyu
UC Berkeley, CA, USA
Google Research, Mountain View, CA, USA

Jelani Nelson
UC Berkeley, CA, USA
Google Research, Mountain View, CA, USA

Tamás Sarlós
Google Research, Mountain View, CA, USA

Uri Stemmer
Tel Aviv University, Israel
Google Research, Herzliya, Israel

Abstract
Composition theorems are general and powerful tools that facilitate privacy accounting across multiple data accesses from per-access privacy bounds. However they often result in weaker bounds compared with end-to-end analysis. Two popular tools that mitigate that are the exponential mechanism (or report noisy max) and the sparse vector technique, generalized in a recent private selection framework by Liu and Talwar (STOC 2019). In this work, we propose a flexible framework of private selection and testing that generalizes the one proposed by Liu and Talwar, supporting a wide range of applications. We apply our framework to solve several fundamental tasks, including query releasing, top-$k$ selection, and stable selection, with improved confidence-accuracy tradeoffs. Additionally, for online settings, we apply our private testing to design a mechanism for adaptive query releasing, which improves the sample complexity dependence on the confidence parameter for the celebrated private multiplicative weights algorithm of Hardt and Rothblum (FOCS 2010).

2012 ACM Subject Classification
Theory of computation → Design and analysis of algorithms

Keywords and phrases differential privacy, sparse vector technique, adaptive data analysis

Digital Object Identifier 10.4230/LIPIcs.ITCS.2023.39

Related Version The arXiv version has a substantially revised introduction, which includes a more accurate survey and comparison with prior works.

Full Version: https://arxiv.org/abs/2211.12063 [6]

Funding Edith Cohen: Partially supported by Israel Science Foundation (grant no. 1595/19).
Xin Lyu: Most of the work was done while interning at Google.
Uri Stemmer: Partially supported by the Israel Science Foundation (grant 1871/19) and by Len Blavatnik and the Blavatnik Family foundation.

1 Introduction

Computations over large datasets often involve multiple intermediate accesses to the data by different algorithms with the final output being some function or aggregate of the outputs of these individual accesses. When we are interested in differential privacy [11], we typically have privacy or stability guarantees for each intermediate access which we would like to use to bound the end-to-end privacy cost.
Generalized Private Selection and Testing with High Confidence

One approach to do that is to use composition theorems [14]. Compositions theorems yield overall privacy cost that scales linearly or sometimes even with square-root dependence on the number of accesses. The disadvantage, however, is that compositions theorems are designed for a scenario where all the intermediate outputs are released and do not benefit in their privacy accounting from the particular way in which the results of the intermediate accesses are aggregated. Powerful tools from the literature for mitigating this generally fall under the two categories of private testing and private selection. Private testing includes the sparse vector technique [12, 32, 20], where each access tests a hypothesis over the data, and only accesses with positive output incur a privacy “charge.” Private selection includes the exponential mechanism [26] that allows to select an approximate best solution (out of a large set of possible solutions) while incurring a privacy cost close to that of estimating the quality of a single solution. These tools have many variations and extensions [24, 22, 30], most prominently, a fairly recent framework by Liu and Talwar [23].

We propose a general template for private selection and testing that is described in Algorithm 1. We view it as a system that provides two functions to users, Selection and Test. The calls to these functions can be interleaved and adaptive (that is, inputs may depend on prior outputs). Algorithm 1 is initialized by randomly sampling an internal parameter $p$ that later serves as a “pass” probability (where the data is not accessed if not “passed”). The distribution of $p$ is controlled by the parameter $\gamma$, where a smaller $\gamma$ pushes the distribution closer to $0$, lowering the privacy cost, and a higher $\gamma$ pushes the distribution closer to $1$, increasing utility. Generally, the choice of $\gamma$ provides a tradeoff between privacy and utility. We will illustrate the tradeoff in several applications later.

The function Selection inputs a parameter $\tau$ and a collection of $k$ mechanisms $\{M_i\}_{i=1}^k$, all with outputs from the same ordered domain. It generates a collection $S$ of outputs and then returns $\text{Best}(S) := \max_{x \in S} x$. The parameter $\tau$ may be set differently in each call to Selection, and its role is to amplify utility by repetition. Intuitively, a larger $\tau$ can compensate for a smaller $p$ as each of the mechanisms $M_i$ produces in expectation $\tau p$ candidates for the Best selection. Surprisingly, as we shall soon see, the privacy accounting allows us to gain from separating $p$ and $\tau$. The function Test provides a basic functionality of hypothesis testing subjected to the “pass” probability $p$. Utility amplification (that also compensates for small pass probability $p$) can be achieved by multiple calls to Test with the same hypothesis.

Meta Privacy Theorems

We establish meta privacy theorems for the system in Algorithm 1. We first provide separate bounds for Selection and Test calls.

**Theorem 1 (Privacy guarantee for Selection).** Suppose that following initialization of Algorithm 1, there are $c \geq 1$ calls to Selection (with possibly varying inputs $\tau, k, \text{ and } \{M_i\}_{i=1}^k$). If the mechanisms fed into Selection in this process satisfy $(\varepsilon, \delta_i)$-DP with $\delta_i \geq 0$, then the list of $c$ outputs satisfies $((2c + \gamma)\varepsilon, \tau \sum_i \delta_i)$-DP.

---

1 For example, the output space of $M_i$ can consist of pairs $(s, q) \in X \times \mathbb{R}$, where $s$ is a solution and $q \in \mathbb{R}$ is a quality score measuring how good $s$ is. We can then order pairs in the decreasing order of $q$, and use Selection to select a solution with the best quality score. We remark that this is also the illustrating example presented in [23].
Theorem 2 (Privacy guarantee for Test). Suppose that following initialization of Algorithm 1, there are repeated calls to Test until \(c \geq 1\) \(\top\) responses are received. If all the mechanisms fed into Test in this process satisfy \((\varepsilon, \delta_i)\)-DP with \(\delta_i \geq 0\), then the collection of outputs satisfies \(((2c + \gamma)\varepsilon, \sum_i \delta_i)\)-DP or \(\left(O\left(\varepsilon \gamma + \varepsilon \sqrt{c \log(1/\delta)}\right), \delta + \sum_i \delta_i\right)\)-DP for all \(\delta > 0\).

The statements of Theorem 1 and Theorem 2 provide separate bounds for the privacy cost of each type of call (Selection or Test). Importantly, our framework allows for interleaved calls where the inputs are adaptive (depending on prior outputs). In this case, the privacy costs simply add up except for the \(\gamma \varepsilon\) factor: If there are \(c_1\) calls to Selection and \(c_2\) calls with \(\top\) output to Test, each with mechanisms that are \((\varepsilon, 0)\)-DP, then the total privacy cost is \(((2c_1 + 2c_2 + \gamma)\varepsilon, 0)\)-DP.

The proofs of the meta-privacy theorems are provided in Section 2. Interestingly, the \(O\left(\varepsilon \sqrt{c \log(1/\delta)}\right)\) dependence in the statement of Theorem 2 does not follow as a direct consequence of advanced composition \([14]\). It requires some delicate calculations and is one of the primary technical contributions of our work.

Algorithm 1 Private Selection and Testing.

Input: A dataset \(D = \{x_1, \ldots, x_n\}\). Parameter \(\gamma \in (0, \infty)\).

1. initialize:
   2. Sample \(p\) from \([0, 1]\) where \(\Pr[p \leq x] = x^\gamma, \forall x \in [0, 1]\).

3. Function Selection(\(\tau, k, M_1, \ldots, M_k\)): // \(M_i\) are mechanisms with outputs from an ordered domain
   4. \(S \leftarrow \emptyset\)
   5. for \(i = 1, \ldots, k\) do
      6. for \(j = 1, \ldots, \tau\) do
         7. \(r_{i,j} \sim \text{Ber}(p)\)
         8. if \(r_{i,j} = 1\) then
            9. \(s_{i,j} \leftarrow M_i(D)\)
            10. \(S \leftarrow S \cup \{s_{i,j}\}\)
   11. return Best\((S)\) // The top value in the ordered set \(S\).

4. Function Test(\(H\)): // \(H\) is a private algorithm with output \(\{\top, \bot\}\)
   5. \(r \sim \text{Ber}(p)\)
   6. if \(r = 1\) then
      7. \(\Gamma \leftarrow H(D)\)
      8. return \(\Gamma\)
   9. else
      10. return \(\bot\)

1.1 Relation to Prior Work

Our framework and its applications can be viewed as both a natural extension of the sparse vector technique (SVT) \([12, 32, 20]\) and a simpler and more versatile alternative to the private selection framework of Liu and Talwar \([23]\).
Comparison with SVT

In a nutshell, SVT allows one to test hypotheses of the form \( f_i(D) \lesssim t \) (i.e., test if \( f_i(D) \) is approximately below the threshold \( t \) up to small additive error) while only incurring privacy loss for “Above” results. We first show that Test calls provide this functionality: To test if \( f_i(D) \lesssim t \) we specify an algorithm \( H \) as follows: \( H \) computes \( f_i(D) = f_i(D) + \text{Lap}(1/\varepsilon) \) and outputs \( \top \) if and only if \( f_i(D) \geq t \). It is easy to see that \( H \) is \((\varepsilon, 0)\)-private. Given the pass parameter \( p \), a single call of Test\((H)\) is insufficient for a reliable response and we therefore amplify the confidence by repetition. For example, when \( p \sim [0, 1] \) it holds that \( \Pr[p \geq \beta] = 1 - \beta \), therefore, if we repeat Test\((H)\) for \( \frac{1}{\beta} \) times and none of these test returns \( \top \), we are very confident that \( f_i(D) \lesssim t \). On the other hand, if \( f_i(D) \leq t - \frac{10}{\varepsilon} \log(1/\beta) \), the probability that we observe a \( \top \) response is at most \( \beta \).

Our meta-privacy proofs for Selection and Test turn out to be natural generalization of the proof of the standard SVT [13]. The role of the “pass” probability \( p \) in our algorithm is very similar to the noisy threshold in SVT, and we prove the privacy property by carefully coupling the executions of Algorithm 1 on two neighboring inputs \( D, D' \) with different \( p \) parameters. See Section 2 for the detail.

In Section 1.5, we further extend our framework to design a private Test procedure, which yields improved confidence-accuracy tradeoffs for SVT. In doing so, we borrow insights from several previous works. This includes the observation that SVT does not necessarily need to re-initialize the noisy threshold in SVT, and we prove the privacy property by carefully coupling the executions of Algorithm 1 on two neighboring inputs \( D, D' \) with different \( p \) parameters. See Section 2 for the detail.

Comparison with Liu-Talwar [23] private selection framework

The hypothesis testing functionality, Test, in our framework does not seem to have an analogue in [23]. The advantages of our Selection procedure compared with [23] are as follows:

- Operationally, [23] implements selection from \( k \) mechanisms by “combining” the \( k \) mechanisms into a single mechanism \( M \), defined as follows. \( M \) chooses a random index \( i \sim [k] \), runs \( M_i \) and outputs its result. Then \( M \) is run for a random number \( T \) times and the best output is returned. A disadvantage is that to ensure utility, we need to set \( T \) to be large enough so that all mechanisms are selected enough times, which increases privacy costs when the base mechanisms \( M_i \)'s only satisfy approximate DP. In contrast, we avoid a random search and consider each mechanism separately, which significantly simplifies the derivation of utility and privacy bounds. This advantage is illustrated in the application to the stable selection task: see Section 1.4.3.

- The privacy analysis in [23] considers one-time private selection calls and is not known to benefit from composition. In contrast, our algorithm (see Theorems 1 and 2) only pays the \( \gamma \varepsilon \) term once (which is possible because we initialize \( p \) once). This tighter composition benefits applications where there are multiple accesses to the data and adaptive accesses. In particular, the tighter composition for Test facilitates our improved sparse vector algorithm.

There is another improvement to [23] by Papernot and Steinke [30], who primarily consider private selection subject to Rényi-DP constraints. They showed that given a private algorithm \( \mathcal{A} \), privately running \( \mathcal{A} \) for \( T \) rounds and selecting the best trial preserves privacy, as long as \( T \) is a random variable drawn from a carefully designed distribution. They established privacy theorems for \( T \) drawn from the Poisson distribution or the truncated negative binomial
distribution. To compare, the Liu-Talwar algorithm shows that when \( T \) obeys a geometric distribution, the privacy is also preserved. Our result implies that when \( T \) is uniform in \( \{0, 1, \ldots, n\} \), the selection is private. However, the algorithm by [30] works by random search, and is subject to the same limitations as we have discussed above: namely, the utility of the algorithm is harder to characterize, and the algorithm is not known to benefit from compositions.

### 1.2 Applications

We overview applications of our framework in Section 1.3, where we demonstrate improvements of results in [23], and in Section 1.4, where we list direct applications of private selection to fundamental tasks in differential privacy, including query releasing, top-\( k \) selection and stable selection.

A separate major contribution of our work is detailed in Section 1.5, where we apply our privacy accounting theorem (Theorem 2) to improve the sparse vector technique (SVT) and through that, improve the private multiplicative weight update algorithm [20] and the sample complexity for adaptive data analysis with linear queries.

### 1.3 Private Selection: Improving better-than-median selection

Liu and Talwar [23] formulated the problem of finding a better-than-median solution of a private algorithm using oracle calls to the algorithm. The problem was also implicitly studied in [19].

**Definition 3 (Private Better-than-median).** Let \( A : X^n \rightarrow Y \times \mathbb{R} \) be an \( (\varepsilon, \delta) \)-DP algorithm, where the output of \( A \) consists of a solution \( y \in Y \) and a score \( s \in \mathbb{R} \).

For \( D \in X^n \) to be a dataset. Let \( s_m = \text{median} \{ s(A(D)) \} \) be the median score one gets when running \( A \) on \( D \). A better-than-median algorithm \( A^* \) with confidence parameter \( \beta \) uses oracle calls to \( A \) and is such that the score of the output of \( A^*(D) \) is larger than \( s_m \) with probability at least \( 1 - \beta \).

The performance of a better-than-median algorithm is measured by computation (number of oracle calls to \( A \)), privacy guarantee, and confidence. Liu and Talwar provided algorithms with the following tradeoffs:

**Theorem 4 ([23]).** For every \( \beta \in (0, 1) \), for \( (\varepsilon, \delta) \)-DP \( A \) with \( \varepsilon \in (0, 1) \) and \( \delta \geq 0 \), there is a better-than-median algorithm \( A^* \) with confidence \( 1 - \beta \) and the following properties.

- In expectation, \( A^* \) makes \( O(\frac{1}{\beta}) \) oracle calls to \( A \).
- \( A^* \) is \( (3\varepsilon, O(\delta/\beta)) \)-differentially private.

Furthermore, for every \( \alpha \in (0, 1) \) and \( \beta, \delta \in (0, 1) \), there is an algorithm \( A^* \) such that:

- With probability 1, \( A^* \) makes \( T \leq \tilde{O}(\frac{1}{\beta^{12+\frac{2\alpha}{\beta}}} \ln(1/\delta)) \) oracle calls to \( A \).
- \( A^* \) is \( ((2 + \alpha)\varepsilon, T\delta + \delta) \)-differentially private.

It was shown in [23] that the privacy guarantee of any better-than-median algorithm is at least \( 2\varepsilon \). Surprisingly, Theorem 4 shows that the failure probability \( \beta \) can be made as small as desired at the cost of (1) additive increase of \( \varepsilon \) in the privacy parameter and (2) number of oracle calls that scales linearly with \( \frac{1}{\beta} \). The privacy cost can be lowered further smoothly to be arbitrarily close to \( 2\varepsilon \) by increasing the number of oracle calls using the parameter \( \alpha \in (0, 1) \).
Our improvement

We design a new algorithm that provides a smooth trade-off between utility and efficiency:

**Theorem 5.** For every $\alpha \in (0, \infty)$ and $\beta \in (0, 1)$, for $(\varepsilon, \delta)$-DP $A$ with $\varepsilon \in (0, 1)$ and $\delta \geq 0$, there is a better-than-median algorithm $A^*$ with confidence $1 - \beta$ that satisfies the following:

- With probability 1, $A^*$ makes at most $T$ oracle calls to $A$, where

$$T = \begin{cases} \left\lceil \frac{2}{2^\gamma} \right\rceil & \text{if } \alpha = 1 \\
\left\lceil 5\left(\frac{2}{2^\gamma}\right)^{1/\alpha} \log(1/\beta) \right\rceil & \text{otherwise} \end{cases}$$

- $A^*$ is $((2 + \alpha)\varepsilon, T\delta)$-differentially private.

Theorem 5 follows using Algorithm 1: Consider the case $\alpha = 1$ and (arbitrary) $\beta \in (0, 1)$. $A^*$ initializes Algorithm 1 with $\gamma = 1$ and calls Selection with arguments $(k = 1, \tau = \left\lceil \frac{2}{2^\gamma} \right\rceil, M_1 = A)$. It follows from Theorem 1 that $A^*$ is $(3\varepsilon, \tau\delta)$-DP. As for the utility, note that Selection runs $A$ for $m \sim \text{Bin}(\tau, p)$ times where $p$ is uniformly random in $[0, 1]$. Equivalently, $m$ is uniformly random in $\{0, 1, \ldots, \tau\}^2$. Therefore, the probability that $A^*$ fails to get a better-than-median solution is at most

$$\frac{1}{\tau + 1} \sum_{m=0}^{\tau} 2^{-m} \leq \frac{2}{\tau + 1} \leq \beta.$$

The privacy-computation tradeoffs by varying $\alpha$, as stated in Theorem 5, are obtained by varying $\gamma$ in the initialization of Algorithm 1. We provide details in Section 3.1. In contrast, in the regime $\alpha \in (0, 1)$, the Liu-Talwar algorithm (Theorem 4) uses estimates of the “quantiles” of the outputs of $A$ and a complex simulation of SVT. Consequently, they only obtain approximate-DP guarantees$^3$ and utilize many more (in the exponent) oracle calls to $A$. Our improvement in the number of oracle calls in the regime $\alpha < 1$ answers an open question posed in [23].

An additional advantage of Theorem 5 is that it applies in the regime $\alpha > 1$ that allow for dramatic improvement in the number of oracle calls when we sacrifice a bit more privacy. For example, with $\alpha = 2$, the number of oracle calls has square-root dependence on $1/\beta$ and with $\alpha = \log(1/\beta)$, the number of oracle calls is only logarithmic in $1/\beta$.

### 1.4 Direct Applications of Private Selection

In the following, we mention several applications of the private selection algorithm. We note that the applications can also be obtained using the Liu-Talwar framework but using Algorithm 1 gives simpler proofs and smaller constants. The primary contribution in this subsection is making the connection between the private selection framework and these concrete applications.

#### 1.4.1 Query Releasing

Consider the task of answering $k$ sensitivity-1 queries $f_1, \ldots, f_k : \mathcal{X}^n \rightarrow \mathbb{R}$. A fundamental and extensively studied [33, 15, 18, 7] problem is to characterize the privacy-accuracy tradeoff for the query releasing task. That is, for a given privacy budget $(\varepsilon, \delta)$, determine the smallest

---

$^2$ As for all $m \in \{0, 1, \ldots, \tau\}$, $E_{p \sim U[0, 1]}(p^m) = \frac{1}{1-p^m} = 1/(\tau + 1)$.

$^3$ Namely, the parameter $\delta$ has to be non-zero in Theorem 4.
error $e$, such that there is an $(\varepsilon, \delta)$-DP algorithm that releases answers to all of the $k$ queries within $\ell_\infty$-error $e$. A naive application of the Gaussian mechanism gives expected $\ell_\infty$-error $O\left(\frac{\sqrt{k \log (1/\delta) \log \log (k)}}{\varepsilon}\right)$, which was improved to $O\left(\frac{\sqrt{k \log (1/\delta) \log \log (k)}}{\varepsilon}\right)$ by Steinke and Ullman [35]. Recently, the question was raised again as an open question [34] and two significant advances were made [18, 7]. We resolve the problem completely by applying the private selection framework together with an algorithm from [18]:

**Theorem 6.** There is a constant $C > 0$ such that the following holds. For every $\varepsilon \in (0,1), \delta \in (0,1/2)$ and $k \in \mathbb{N}$, there is an $(\varepsilon, \delta)$-DP algorithm that answers $k$ given sensitivity-1 queries $f_1, \ldots, f_k$, such that

$$
\Pr_{\hat{f}} \left[ \|f - \hat{f}\|_\infty \leq C \frac{\sqrt{k \log (1/\delta) + \log (1/\delta)}}{\varepsilon} \right] = 1,
$$

where $\hat{f} := (\hat{f}_1, \ldots, \hat{f}_k)$ denotes the responses released by the algorithm.

**Remark 7.** The $\frac{\log (1/\delta)}{\varepsilon}$ term in the error is necessary if one desires the answers to be accurate with probability one. There is a lower bound even for the case $k = 1^4$.

Theorem 6 is tight, and improves on the two incomparable previous works [18, 7].

- [7] shows how to achieve the aforementioned accuracy with probability 1. However, their algorithm only works for $\delta > 2^{-\frac{k}{\log^2(k)}}$.
- [18] shows an algorithm for the full range of $\delta \in (0,1/2)$. However, their algorithm only promises to release an answer vector $\hat{f}$ that is accurate (in $\ell_\infty$ sense) with probability $1 - \frac{4}{\log (k)}$. It was posed as an open question in [18] whether one can release accurate responses (within $\ell_\infty$ distance) with probability one. Theorem 6 answers this question affirmatively.

**Proof intuition**

In [18], it was shown how to $(\varepsilon, \delta^2)$-privately release a vector $(\hat{f}_1, \ldots, \hat{f}_k)$ such that, with probability at least $\frac{1}{2}$, one has

$$
\|f - \hat{f}\|_\infty \leq O\left(\frac{\sqrt{k \log (1/\delta)}}{\varepsilon}\right).
$$

In the following, we use $A$ to denote the GKM algorithm with privacy parameters $(\varepsilon, \delta^2)$.

The high level idea is to use the better-than-median algorithm of Theorem 5 to amplify the success probability. Consider applying Theorem 5 on top of $A$ with $\alpha = 1$ and $\beta = \delta$. Theorem 5 shows that the number of oracle calls to $A$ is bounded by $O(1/\delta)$. Then, it follows that the resulting algorithm is $(O(\varepsilon), O(\delta))$-differentially private. Moreover, the algorithm releases an accurate vector with probability $1 - \delta$. We can further improve the success probability to 1 by non-privately correcting erroneous vectors. Since we do the non-private correction with probability at most $\frac{1}{2}$, the result algorithm is still $(O(\varepsilon), O(\delta))$-DP. See Section 3.2 for the detail.

---

[^4]: To see this lower bound, let $A$ be an arbitrary query-releasing mechanism. We construct a sensitivity-1 function $f$ and two private inputs $D^0, D^1$ of distance $\frac{\log (1/\delta)}{\varepsilon}$ such that $f(D^0) - f(D^1) = \frac{\log (1/\delta)}{\varepsilon}$. Since $A$ is $(\varepsilon, \delta)$-DP, the supports of $A(f, D^0)$ and $A(f, D^1)$ overlap, which means that there is $v \in \mathbb{R}$ and $D \in \{D^0, D^1\}$ such that $|v - f(D)| \geq \frac{\log (1/\delta)}{\varepsilon}$ but $\Pr[A(f, D) = v] > 0$. 

---

ITCS 2023
1.4.2 Top-k Selection

We consider optimization over a finite set, which is another fundamental task in DP. Suppose there are \( m \) candidates. Each of them is associated with a sensitivity-1 score function \( f_1, \ldots, f_m : X^n \to \mathbb{R} \). We want to select \( k \) candidates from the \( m \) options with largest scores. This is a well-known and fundamental task that has been studied extensively (see, e.g., [26, 13, 25, 9, 8, 31]).

To measure the utility of a selection algorithm, a common criterion is the suboptimality gap. Suppose the algorithm releases a subset \( S \subseteq [m] \) of candidates. The suboptimality gap of \( S \) is defined as

\[
\text{Gap}(S) := \max_{i \not\in S} \{ f_i(D) \} - \min_{j \in S} \{ f_j(D) \}.
\]

Namely, this is the difference between the \textit{best} option outside \( S \) and the \textit{worst} option inside \( S \).

The standard algorithm for top-\( k \) applies the exponential mechanism sequentially for \( k \) times. If we aim for \((\varepsilon, \delta)\)-DP for the final algorithm, each round of the exponential mechanism needs to be \((\varepsilon \sqrt{k \log(1/\delta)}, 0)\)-DP. Let \( S \) denote the output of this algorithm. The standard calculation shows that

\[
\Pr \left[ \text{Gap}(S) \geq \frac{\sqrt{k \log(1/\delta)}}{\varepsilon} \cdot \log(m/\beta) \right] \leq \beta.
\]

Namely, with probability \( 1 - \beta \), the suboptimality gap is \( \frac{\sqrt{k \log(1/\delta)}}{\varepsilon} \cdot \log(m/\beta) \). If one is interested in very high confidence \( \beta \leq m^{-\omega(1)} \), this bound on the gap can be very large. We offer an algorithm where the suboptimality gap has improved dependency on \( 1/\beta \).

\begin{theorem}
There is an absolute constant \( C > 0 \) such that the following holds. For every \( \varepsilon, \delta, \beta \in (0, 1) \), there is an \((\varepsilon, \delta)\)-DP algorithm for top-\( k \) selection satisfying the following. Let \( S \) denote the output of the algorithm. Then

\[
\Pr \left[ \text{Gap}(S) \geq C \frac{\sqrt{k \log(1/\delta)}}{\varepsilon} \cdot \log(m) + C \frac{\log(1/\beta)}{\varepsilon} \right] \leq \beta.
\]
\end{theorem}

While we cannot avoid paying \( \log(1/\beta) \) completely, we only pay \( \log(1/\beta) \) once instead of \( \sqrt{k \log(1/\delta)} \) times as was in the standard algorithm.

The proof of Theorem 8 uses a similar strategy as that of Theorem 6. Suppose we iteratively run the exponential mechanism for \( k \) rounds to get \( k \) candidates \( S \). With probability at least \( \frac{1}{2} \) we have that \( \text{Gap}(S) \leq O(\frac{\sqrt{k \log(1/\delta)}}{\varepsilon} \cdot \log(m)) \). We can use Theorem 5 on top of the naive algorithm to boost the success probability from \( \frac{1}{2} \) to \( 1 - \beta \).

1.4.3 Stable Selection

The private selection algorithm also has implications to stable selections. Generally, there is a lower bound saying that the \( \Omega(\frac{1}{\varepsilon} \log(m)) \) suboptimality gap is necessary when we select from \( m \) candidates with sensitivity-1 score functions. However, when there is a “structure” within the candidate family, we can often do better. We show two representative examples in the following.
Choosing Mechanism

Let $F$ be a family of sensitivity-1 functions. Say that $F$ is $k$-bounded, if for any two neighboring datasets $D, D'$, it holds that $\sum_{f \in F} |f(D) - f(D')| \leq k$. For $k$-bounded function families, we can do private selection with suboptimality gap $O(\frac{1}{\varepsilon} \log(k))$.

\textbf{Theorem 9.} Suppose $F$ is a $k$-bounded function family. For every $\varepsilon, \delta \in (0, 1)$, there is an $(\varepsilon, \delta)$-DP algorithm $A$ that holds a dataset $D$ and selects a function $f \in F$ such that, with probability at least $1 - \beta$,

$$f(D) \geq \max_{f^* \in F} \{f^*(D)\} - O(\frac{1}{\varepsilon} \log(k/\delta \beta)).$$

Theorem 9 generalizes several previous mechanisms for this setting [2, 17], which all require additional assumptions on the family $F$.

Stable Selection

In some settings, there is a noticeable quality gap between the best solution in $F$ and the $(k + 1)$-th best solution. In this case, we can also obtain a private selection algorithm with improved suboptimality gap. In more detail, let $F$ be the function family. For any given dataset $D$, sort functions in $F$ by the decreasing order of $f(D)$. Namely, write $F = \{f_1, f_2, \ldots, f_m\}$ so that $f_1(D) \geq f_2(D) \geq \cdots \geq f_m(D)$. Define the $k$-th gap of $D$ on $F$ as $G_k(D, F) = f_1(D) - f_{k+1}(D)$.

\textbf{Theorem 10.} For any $k \geq 1$ and $\varepsilon, \delta, \beta \in (0, 1)$, there is an $(\varepsilon, \delta)$-DP algorithm $A$ that holds a dataset $D$ and achieves the following. For any 1-Lipschitz function family $F$, suppose $G_k(D, F) \geq \frac{10}{\varepsilon} \log(k/\delta \beta)$. Then, with probability at least $1 - \beta$, $A$ selects a function $f \in F$ such that

$$f(D) \geq \max_{f^* \in F} \{f^*(D)\} - O(\frac{1}{\varepsilon} \log(k/\delta \beta)).$$

There is a known algorithm for this task with asymptotically the same performance (see [3, 4] and references therein). However, our algorithm and its analysis are very different. In particular, we demonstrate how easily this problem can be solved under the private selection framework.

The proofs of Theorems 9 and 10 are deferred to Section 3.3 and the full version of the paper.

1.5 Improved Sparse Vector Technique with Applications

In Section 1.4, we have shown that one can use the private selection framework to improve the confidence-accuracy tradeoff for query releasing and top-$k$ selection. However, note that these two applications work in the offline setting: in the query releasing and top-$k$ selection tasks, we are given the $k$ queries and the $m$ candidates in advance, respectively.

It is natural to ask if similar improvements can be obtained in the online setting. In the following, we demonstrate this is possible by designing a modification of the well-known sparse vector technique, which offers an improved confidence-accuracy tradeoff. We further apply our SVT to improve the private multiplicative weight algorithm [20] and the sample complexity of adaptive data analysis.
1.5.1 Review of Sparse Vector Technique and its Applications

Roughly speaking, the sparse vector technique provides an algorithm to answer an *unbounded number* of queries of the form \( f_i(D) \lesssim t \), while one only incurs privacy loss for queries that are above the thresholds. See Section 1.5.2 for a more technical and precise description.

Application to private multiplicative weights

SVT is the main subroutine in the celebrated private multiplicative weights algorithm [20], which can be used to answer a large number of linear queries subject to privacy constraints. We recall the problem setup first. Suppose there is a universe \( \mathcal{X} \) of size \(|\mathcal{X}|\). We want to design an algorithm \( \mathcal{A} \) that receives a private input \( D = (x_1, \ldots, x_n) \in \mathcal{X}^n \) consisting of \( n \) data points from \( \mathcal{X} \), and answers \( m \) queries from an adversary \( B \). Each query is specified by a function \( f_i : \mathcal{X}^n \to [0, 1] \), and the algorithm needs to report an estimation \( \hat{f}_i \) for \( \mathbb{E}_{j \sim [n]}[f_j(x_j)] \). The error of the algorithm is defined as \( \max_{i \in [m]} \left| \hat{f}_i - \mathbb{E}_{j \sim [n]}[f_j(x_j)] \right| \).

The private multiplicative weight update (MWU) algorithm allows one to privately answer \( m \) linear queries, while the error grows only logarithmically with \( m \), as shown in the following theorem.

**Theorem 11** ([20], see also [13]). For every \( \varepsilon, \delta \in (0, 1) \), there is an \((\varepsilon, \delta)\)-DP algorithm that answers \( m \) adaptively chosen linear queries with the following utility guarantee. Let \((\hat{f}_i)_{i \in [m]}\) denote the outputs of \( \mathcal{A} \). For every \( \beta \in (0, 1) \), with probability \( 1 - \beta \) we have that \( \max_{i \in [m]} \left| \hat{f}_i - \mathbb{E}_{j \sim [n]}[f_j(x_j)] \right| \leq \alpha = \alpha(\beta) \), where

\[
\alpha(\beta)^2 \leq O \left( \frac{\sqrt{\log |\mathcal{X}| \log(1/\beta)} \log(m/\beta)}{\varepsilon} \right).
\]

Theorem 11 can be used to perform adaptive data analysis with linear queries. We recall the setup: given the universe \( \mathcal{X} \), there is a distribution \( \mathcal{D} \) over \( \mathcal{X} \). An adaptive data analysis algorithm receives as input \( n \) samples from \( \mathcal{D} \), denoted by \( S \sim \mathcal{D}^n \). It needs to adaptively answer \( m \) linear queries, where each query is specified by a function \( f_i : \mathcal{X} \to [0, 1] \) and the algorithm needs to output an estimation of \( \mathbb{E}_{x \sim \mathcal{D}}[f(x)] \). An algorithm is called \((\alpha, \beta, m)\)-accurate, if it can answer \( m \) (adaptive) linear queries within additive error \( \alpha \) with probability at least \( 1 - \beta \). We consider the sample complexity of the problem. That is, fixing \( \mathcal{X}, m, \alpha, \beta \), we want to find out the smallest \( n \) such that there exists an \((\alpha, \beta, m)\)-accurate algorithm that only uses \( n \) samples from \( \mathcal{D} \).

The generalization property of DP [10, 1] shows that, if we design an algorithm to output estimations of \( \mathbb{E}_{x \sim \mathcal{S}}[f_i(x)] \) in a privacy-preserving manner, then \( \mathbb{E}_{x \sim \mathcal{S}}[f_i(x)] \) is a good approximation of \( \mathbb{E}_{x \sim \mathcal{D}}[f_i(x)] \). Combining the privacy-preserving algorithm by Theorem 11 with the generalization property, one obtains the following well-known sample complexity upper bound.

**Theorem 12** ([1]). For every finite universe \( \mathcal{X} \), every \( m \geq 1 \) and \( \alpha, \beta \in (0, 1) \), there is an \((\alpha, \beta, m)\)-accurate adaptive data analysis algorithm with sample complexity

\[
n \leq O \left( \frac{\sqrt{\log |\mathcal{X}| \log(1/\beta)} \log(m/\beta)}{\alpha^3} \right).
\]

Our improvement

Using our improved sparse vector algorithm, we obtain the following improved private multiplicative weight update algorithm.
Theorem 13. There is an absolute constant $C > 0$ such that the following is true. For every $\varepsilon, \delta \in (0, 1)$, $m \in \mathbb{N}$ and $\beta \geq 2^{-m}$, there exists

$$\alpha \leq C \cdot \left( \frac{\sqrt{\log |X| \log(1/\delta)} \log(m)}{n \varepsilon \cdot \alpha} + \frac{\log(1/\beta)}{n \varepsilon} \right)$$

and an $(\varepsilon, \delta)$-DP algorithm that answers $m$ adaptively chosen linear queries such that, with probability at least $1 - \beta$, we have $\max_{i \in [m]} |\tilde{f}_i - E_{j \in [n]} [f_i(x_j)]| \leq \alpha$.

We get the following improved sample complexity for adaptive data analysis as a corollary of Theorem 13 and the generalization property by [1].

Theorem 14. For every finite universe $X$, every $m \geq 1$ and $\alpha \in (0, 1), \beta \geq 2^{-m}$, there is an $(\alpha, \beta, m)$-accurate adaptive data analysis algorithm with sample complexity

$$n \leq O \left( \min \left( \frac{\sqrt{\log |X| \log(1/\beta)} \log |X|}{\alpha^3}, \frac{\log |X|}{\alpha^4} \right) \log m + \frac{\log(1/\beta)}{\alpha^2} \right).$$

We compare Theorem 14 with Theorem 12. Overall, the sample complexity given by Theorem 14 scales at most linearly with $\log(1/\beta)$, and we manage to “decouple” the dependence on $\log(1/\beta)$ from that on other parameters. For the non-adaptive setting, it is known (by the Chernoff bound) that the optimal sample complexity is $\Theta(\frac{\log m}{\alpha^2})$. Therefore, in the high confidence regime (in particular, when $\log(1/\beta) > \frac{\log |X|}{\alpha^2} \log m$), the bound given by Theorem 14 asymptotically matches the non-adaptive bound.

1.5.2 Intuition

Now we discuss the proof idea of our improved sparse vector technique.

Review of the standard SVT

Let us briefly review how the standard SVT works. Let $X$ be the private input, $t \in \mathbb{R}$ be a threshold, and $f_1, \ldots, f_m$ be a list of sensitivity-1 queries. The sparse vector algorithm works by adding independent Laplace noises $\text{Lap}(1/\varepsilon)$ to the threshold and each query function. Then, it outputs the first index $i$ such that the (noisy) query value $\tilde{f}_i(X)$ is larger than the (noisy) threshold $\tilde{t}$.

The standard approach to argue the utility of the algorithm works as follows. First, with probability at least $1 - \beta$, all the Laplace noises sampled in the SVT algorithm are bounded by $\frac{1}{\beta} \log(m/\beta)$. Conditioning on this event and letting $i^*$ be the index returned by the algorithm, we have that $f_i(X) \geq t - \frac{2}{\beta} \log(m/\beta)$ and $f_j(X) \leq t + \frac{2}{\beta} \log(m/\beta)$ for every $j < i^*$. Therefore, we conclude that the algorithm is “approximately accurate” within error $O(\frac{1}{\beta} \log(m/\beta))$ with probability $1 - \beta$.

In many scenarios (including the private MWU algorithm), one usually wants to repeatedly run SVT for $k$ rounds, to identify $k$ meaningful queries. Suppose we desire the final algorithm to be $(\varepsilon, \delta)$-private. In each round of SVT, we need to set the privacy budget as $\varepsilon' = \frac{\varepsilon}{\sqrt{k \log(1/\delta)}}$.

Doing the analysis above, the error bound becomes $\frac{1}{\beta} \sqrt{\frac{k \log(1/\delta)}{\log(m/\beta)}}$. This is somewhat unsatisfactory, as we need to pay the $\log(1/\beta)$ term “$O(\sqrt{k})$” times.

\[\] However, we remark that the $\sqrt{k \log(m)}$ term is necessary (i.e., there is a lower bound of $O(\sqrt{k \log(m)})$. See, e.g. [5, 33, 36]).
Our improvement

In the following, we show a modification of the sparse vector algorithm with better confidence-accuracy trade-off. In particular, our algorithm is accurate within error $O(\frac{1}{2} \sqrt{k \log(1/\delta) \log(m) + \frac{\log(1/\beta)}{\delta}})$ with probability $1 - \beta$. To compare, the standard SVT has error $O(\frac{1}{2} \sqrt{k \log(1/\delta) \log(m/\beta)})$. See the full version of the paper for the formal description of our new algorithm. We explain the intuition and the main technical challenges below.

Recall that we want to run SVT for $k$ rounds, and desire the final privacy guarantee to be $(\varepsilon, \delta)$-DP. Given this constraint, each round of SVT has to be $(\varepsilon', 0)$-DP for $\varepsilon' = \frac{\varepsilon}{\sqrt{k \log(1/\delta)}}$.

To illustrate the idea, suppose for now that we do not need to add a noise to the threshold. (Namely, let us suppose that $\hat{t} = t$ always holds.) Now, given a query $f_i(X)$, the sparse vector algorithm first obtains a noisy estimation $\hat{f}_i(X) = f_i(X) + \text{Lap}(1/\varepsilon')$, and then compares $\hat{f}_i(X)$ against $\hat{t}$. If the comparison result is $\hat{f}_i(X) \leq \hat{t}$, then we are $(1 - \beta)$-confident that $f_i(X) \leq t + \frac{2}{\beta} \log(1/\beta)$. Observe that the error gap $\frac{2}{\beta} \log(1/\beta)$ keeps increasing as we desire a higher and higher confidence.

We can amplify confidence more economically by repetition. Let $\tau = \log(1/\beta)$. Consider drawing $\tau$ independent noisy estimations of $f_i(X)$. Namely, we calculate $\hat{f}_i^{(j)}(X) = f_i(X) + \text{Lap}(1/\varepsilon')$ for every $j \in [\tau]$. Suppose that all of these noisy estimations are below $\hat{t}$. Then we are $(1 - \beta)$-confident that $f_i(X) \leq \hat{t}$. Here, somewhat magically, we get the higher confidence without compromising privacy, as the SVT algorithm only incurs a privacy loss when one observes an “Above-Threshold” result.

However, the situation becomes more complicated when we get the “Above-Threshold” result. If there is one estimation $\hat{f}_i^{(j)}(X)$ exceeding $\hat{t}$, we cannot immediately declare that $f_i(X) \geq \hat{t}$. Indeed, for every fixed $f_i(X)$ and $\hat{t}$, as $\tau$ tends to infinity, with probability one, we will eventually see a noisy estimation that is above $\hat{t}$. Hence, the larger $\tau$ is, the less confident we are about the conclusion $f_i(X) \geq \hat{t}$. To ensure that we can be equally confident in the “Above-Threshold” case, we use the following strategy. We set a new threshold $\hat{t}_{\text{lower}} = \hat{t} - \frac{6 \log(m)}{\beta \varepsilon'}$. Then we test if $f_i(X) \geq \hat{t}_{\text{lower}}$. Similarly, we amplify confidence by repetition. Namely, we obtain $\tau$ independent estimations of $f_i(X)$, and compare them against $\hat{t}_{\text{lower}}$. We pass the test if and only if all the $\tau$ estimations are above $\hat{t}_{\text{lower}}$. If we pass the test, then we are very confident that $f_i(X) \geq \hat{t}_{\text{lower}}$. If we fail the test, we switch back to test if $f_i(X) \leq \hat{t}$ (using another $\tau$ independent estimations of $f_i(X)$). In general, we alternate between two tests $f_i(X) \leq \hat{t}$ and $f_i(X) \geq \hat{t}_{\text{lower}}$, until we pass one of them.

Two challenges

To implement the idea, there are two remaining issues.

- In the standard SVT algorithm, we only pay for each “Above-Threshold” answer, and every such answer can identify a meaningful query for us. However, in our modified version of SVT, we will repeatedly test $f_i(X) \leq \hat{t}$ and $f_i(X) \geq \hat{t}_{\text{lower}}$ until we pass one of them. Each time we fail a test, we need to pay the privacy loss. How do we ensure that we can use the vast majority of our “privacy budget” to answer meaningful queries?

- In the argument above, we ignored the issue that $\hat{t}$ is only a noisy version of $t$. Indeed, if we construct $\hat{t}$ by $\hat{t} = t + \text{Lap}(1/\varepsilon')$, we are only $(1 - \beta)$-confident that $|\hat{t} - t| \leq \frac{1}{\beta} \log(1/\beta)$. Therefore, there is a noticeable chance (e.g., with probability $10\beta$) that we will start with an erroneous noisy threshold $\hat{t}$ (e.g., $|\hat{t} - t| \geq \frac{1}{\beta} \log(\frac{1}{10\beta})$). If this does happen, then the “alternate testing” algorithm does not make much sense.
We resolve the first issue by allowing a gap of \( \frac{9}{\epsilon'} \log(m) \) between \( \hat{t}_{\text{lower}} \) and \( \hat{t} \). Then, for any query \( f_i(X) \), at least one of the following is true:

\[
f_i(X) \leq \hat{t} - \frac{3}{\epsilon'} \log(m), \quad \text{or} \quad f_i(X) \geq \hat{t}_{\text{lower}} + \frac{3}{\epsilon'} \log(m).
\]

We assume \( \tau = \log(1/\beta) \leq m \). In this case, if \( f_i(X) \leq \hat{t} - \frac{3}{\epsilon'} \log(m) \), we can pass the test \( f_i(X) \leq \hat{t} \) with probability at least \( 1 - \frac{1}{m\tau} \). If \( f_i(X) \leq \hat{t} - \frac{3}{\epsilon'} \log(m) \) and we do fail the test \( f_i(X) \leq \hat{t} \), then we say that a “mistake” happens and one round of SVT is “wasted”. Similarly, we will say one round of SVT is wasted if \( f_i(X) \geq \hat{t}_{\text{lower}} + \frac{3}{\epsilon'} \log(m) \) but we fail the test \( f_i(X) \geq \hat{t}_{\text{lower}} \). Since the probability of making a mistake is small (i.e., the probability is at most \( \frac{1}{m\tau} \)), we can show that if we run our version of SVT for \( k \) rounds to handle \( m \) queries, then with probability \( 1 - \beta \), at most \( \frac{\log(1/\beta)}{\log(m)} \) rounds are wasted. Therefore, we can still use the algorithm to identify \( k - \frac{\log(1/\beta)}{\log(m)} \) meaningful queries.

We resolve the second issue by noting that there is a version of SVT that only requires one to noisify the threshold once (see, e.g., [24]). Here, the crucial point is that when we only add noise to the threshold once, we can add a much smaller noise to ensure privacy. However, the result from [24] only considers pure-DP case, and the privacy loss scales linearly with \( k\epsilon \) in their analysis. We manage to prove Theorem 2, which shows that the privacy scales proportionally with \( \sqrt{k}\epsilon \) if we relax the requirement to approximate DP.

To be more precise, we explain the utility improvement quantitatively. In the standard SVT algorithm, we need to re-initialize the noisy threshold before starting each of the \( k \) rounds, and we draw noises from the distribution \( \text{Lap}(1/\epsilon') \) to achieve privacy. With our new composition theorem, we only need to add one noise drawn from \( \text{Lap}(1/\epsilon) \). Recall that \( \frac{1}{\epsilon'} \approx \sqrt{\frac{\log(1/\beta)}{\epsilon}} \). The latter noise is considerably smaller than the former one.

In the formal proof, we design our algorithm under the framework of Algorithm 1, using the Test procedure as the main subroutine, which turns out to be more amenable to mathematical manipulations. In the discussion above, we chose to discuss our algorithmic idea under the standard SVT framework, because SVT is arguably more well-known, and it might help readers gain the intuition better.

The possibility of not noisifying threshold

We also note that there are variants of SVT that do not add noise to the threshold at all (see, e.g., [20, 21]). It is tempting to apply those variants to address the second issue. However, looking into the privacy proofs for those variants, one can note that their \( \delta \) parameter has to be at least \( 2^{-k} \), where \( k \) is the number of rounds that we run SVT (i.e., the number of \( \top \) responses we can get before halting the algorithm). In most applications, requiring \( \delta \) to be at least \( 2^{-k} \) would not be an issue. However, in the application to private multiplicative weights and data analysis, we are also interested in the case where \( \delta \ll 2^{-k} \). In this case, adding a noise to the threshold and using Theorem 2 to do the privacy accounting give us the best possible result.

More precisely, if we use a version of SVT that does not noisify the threshold, we will end up with weaker versions of Theorems 13 and 14. Specifically, the error bound in Theorem 13 degrades to

\[
O \left( \frac{\sqrt{\log |X| \log(1/\delta) \log(m)}}{n \epsilon \alpha} + \frac{\log(1/\delta) \log(m)}{n \epsilon} \right).
\]
and the sample complexity in Theorem 14 degrades to
\[ n \leq O \left( \min \left( \frac{\sqrt{\log |\mathcal{X}| \log (1/\beta)}}{\alpha^4}, \frac{\log |\mathcal{X}|}{\alpha^4} \right) \log m + \frac{\log (1/\beta)}{\alpha^2} \right). \]

## 2 Proof of Privacy Theorems

### 2.1 Proof of Theorem 1

#### Proof.

We first consider the case that all the mechanisms \( M_i \)'s satisfy pure-DP (i.e., \( \delta_i = 0 \) for all \( i \)). At the end the proof, we explain how to deal with candidates with approximate DP property. Let \( P_\gamma \) denote the CDF for the parameter \( p \). Namely, \( P_\gamma(v) = \Pr[x \leq v] = v^\gamma \).

Fix \( c \geq 1 \). Let \( \mathcal{B} \) be an adversary that adaptively calls \( \text{Selection} \) for \( c \) rounds. Denote Algorithm 1 as \( \mathcal{A} \). For each \( p \in [0, 1] \), let \( \mathcal{A}^p \) denote Algorithm 1 conditioning on it having sampled \( p \) in the initialization step.

In one round, suppose that the \( \mathcal{B} \) calls \( \text{Selection} \) with \((\tau, k, M_1, \ldots, M_k)\). By duplicating each mechanism for \( \tau \) times, we can assume without loss of generality that \( \tau = 1 \). Let \((i, s) \in [k] \times \mathcal{Y} \) be a potential outcome. Let \( D \) and \( D' \) be two neighboring datasets. Let \( \mathcal{A}^p(D, (M_1, \ldots, M_k)) \) denote the output of \( \mathcal{A}^p \) given input \( D \) and query \((M_1, \ldots, M_k)\). We prove

\[
\Pr[\mathcal{A}^p(D, (M_1, \ldots, M_k)) = (i, s)] \leq e^{2\epsilon} \Pr[\mathcal{A}^{p/\epsilon}(D', (M_1, \ldots, M_k)) = (i, s)].
\]

Note that \( \mathcal{A} \) on this \( \text{Selection} \) call returns \((i, s)\), if and only if both of the following events hold:

- Event \( \mathcal{E}_1 \): \( r_{i,1} = 1 \) and \( M_i \) outputs \( s \).
- Event \( \mathcal{E}_2 \): For every \( i' \neq i \), either \( r_{i',1} = 0 \) or \( M_{i'} \) outputs a solution inferior than \( s \).

Now, observe that

\[
\Pr[\mathcal{E}_1 \mid \mathcal{A}^p(D)] = p \cdot \Pr[M_i(D) = s] \\
\leq e^{\epsilon} p \Pr[M_i(D') = s] \\
\leq e^{2\epsilon} \frac{p}{e^\epsilon} \Pr[M_i(D') = s] \\
= e^{2\epsilon} \Pr[\mathcal{E}_1 \mid \mathcal{A}^{p/\epsilon}(D')].
\]

(1)

For every \( i' \neq i \), we have

\[
\Pr[r_{i',1} = 1 \land M_{i'} \text{ outputs a solution better than } s \mid \mathcal{A}^p(D)] \\
= p \cdot \Pr[M_{i'}(D) \text{ outputs a solution better than } s] \\
\geq \frac{p}{e^\epsilon} \cdot \Pr[M_{i'}(D') \text{ outputs a solution better than } s] \\
= \Pr[r_{i',1} = 1 \land M_{i'} \text{ outputs a solution better than } s \mid \mathcal{A}^{p/\epsilon}(D')].
\]

Therefore,

\[
\Pr[r_{i,1} = 0 \lor M_i \text{ outputs a solution worse than } s \mid \mathcal{A}^p(D)] \\
\leq \Pr[r_{i,1} = 0 \lor M_i \text{ outputs a solution worse than } s \mid \mathcal{A}^{p/\epsilon}(D')].
\]

Consequently,

\[
\Pr[\mathcal{E}_2 \mid \mathcal{A}^p(D)] \leq \Pr[\mathcal{E}_2 \mid \mathcal{A}^{p/\epsilon}(D')].
\]

(2)
Combining (1) and (2) yields that
\[ \Pr[A^p(D, (M_1, \ldots, M_k)) = (i, s)] \leq e^{2\varepsilon} \Pr[A^{p/\varepsilon'}(D', (M_1, \ldots, M_k)) = (i, s)]. \]

Note that this inequality holds for all \((M_1, \ldots, M_k)\), independently of the private input. Now, suppose \(B\) interacts with \(A^p(D)\) (or \(A^{p/\varepsilon}(D')\)) for \(c\) rounds. We can apply the argument above for the \(c\) rounds separately. Let \(IT(B : A)\) denote the random variable recording the interaction transcript between \(B\) and \(A\). For every possible collection \(E\) of outcomes, we have
\[ \Pr[IT(B : A^p(D)) \in E] \leq e^{2\varepsilon} \Pr[IT(B : A^{p/\varepsilon}(D')) \in E]. \]

Finally, we have
\[
\begin{align*}
\Pr[IT(B : A(D)) \in E] &= \int_0^1 \Pr[IT(B : A^p(D)) \in E] dP(p) \\
&\leq \int_0^1 e^{2\varepsilon} \Pr[IT(B : A^{p/\varepsilon}(D')) \in E] \cdot (\gamma p^{\gamma - 1}) \cdot dp \\
&\leq \int e^\varepsilon \cdot \Pr[IT(B : A^{p/\varepsilon}(D')) \in E] \cdot (\gamma(p/e)^{\gamma - 1}) \cdot e^{(\gamma - 1)\varepsilon} \cdot d(p/e) \\
&\leq \int e^{(2c + \gamma)e} \Pr[IT(B : A^{p/\varepsilon}(D')) \in E] \cdot dP(p/e) \\
&= e^{(2c + \gamma)e} \Pr[IT(B : A(D')) \in E].
\end{align*}
\]

This completes the proof.

Now, we consider the case that each mechanism \(M_i\) is \((\varepsilon, \delta_i)\)-DP with \(\delta_i \geq 0\). For two neighboring inputs \(D, D'\), we can decompose \(M_i(D) = (1 - \delta_i)N_i(D) + \delta_iE_i(D)\) and \(M_i(D') = (1 - \delta_i)N_i(D') + \delta_iE_i(D')\), where \(N_i(D)\) and \(N_i(D')\) are \((\varepsilon, 0)\)-indistinguishable and \(E_i(D), E_i(D')\) are arbitrary. Then, each time we run \(M_i(D)\), we can first sample from \((N_i(D), E_i(D))\) with probability \((1 - \delta, \delta)\), and run the chosen mechanism. It follows that with probability at least \(1 - \tau\delta_i\), all the executions involve only the \(N_i\)-part. We union-bound over \(i\), and conclude that with probability at least \(\tau \sum_i \delta_i\), all the executions involve only the \(N_i\)-part. Conditioning on this event, the argument above shows that the outputs of the algorithm on \(D, D'\) are \(((2c + \gamma)e, 0)\)-DP. It follows that the original system is \(((2c + \gamma)e, \tau \sum \delta_i)\)-DP.

### 2.2 Proof of Theorem 2

In this section, we prove Theorem 2. To ease the presentation, we will assume that all the mechanisms fed into \(\text{Test}\) satisfy pure-DP. Having proved for this case, the proof for approximate-DP mechanisms can be verified using similar arguments as in the proof of Theorem 1.

#### 2.2.1 Preliminaries

We will use the Rényi Differential Privacy [27] framework to prove Theorem 2. Recall the definition. For two distributions \(P, Q\) on the universe \(X\) and a real \(\alpha \in (1, \infty)\), we define the \(\alpha\)-order Rényi divergence of \(P\) from \(Q\) as
\[
D_\alpha(P\|Q) := \frac{1}{\alpha - 1} \log \left( \sum_{x \in X} P(x) \left( \frac{P(x)}{Q(x)} \right)^{\alpha - 1} \right).
\]
The max-divergence is defined as
\[ D_\infty(P\|Q) := \log(\max_x \{P(x)/Q(x)\}). \]

Since we will defer most of the technical manipulations to the full version of the paper, we don’t discuss properties of divergences here.

### 2.2.2 Reduction to an Asymmetrical Coin Game

We use \( \mathcal{A} \) to denote Algorithm 1. For \( p \in [0, 1] \), let \( \mathcal{A}^p \) denote the algorithm \( \mathcal{A} \) conditioning on it having sampled \( p \) in the initialization step. Fix two neighboring dataset \( S, S' \). Let \( IT(\mathcal{B} : \mathcal{A}^p(S)) \) denote the random variable recording the interaction between the two randomized systems \( \mathcal{B} \) and \( \mathcal{A} \). Let \( \mathcal{B} \) be an adversary that interacts with \( \mathcal{A} \) by calling \( Test \). Fix \( p \in [0, 1] \), we first compare \( IT(\mathcal{B} : \mathcal{A}^p(S)) \) with \( IT(\mathcal{B} : \mathcal{A}^{p/\varepsilon}(S')) \). It turns out that the two interactions can be captured by the following experimental game, which we call the “0-favored” coin flipping game. Our formulation of the coin-flipping game is inspired by several coin-flipping mechanisms appeared in previous works [19, 29, 28, 22].

**Algorithm 2** The 0-favored coin-flipping mechanism.

**Input:** An input bit \( b \in \{0, 1\} \), a privacy parameter \( \varepsilon \in (0, 1) \), an integer \( k \geq 1 \).

1. **initialize:**
2. \( c \leftarrow 0 \)
3. **while** \( c < k \) **do**
4. \( (p, q) \in [0, 1]^2, \text{promised that } 0 \leq q \leq p \leq e^\varepsilon q \text{ and } (1 - q) \leq e^\varepsilon (1 - p) \)
5. **if** \( b = 0 \) **then**
6. Sample \( r \sim \text{Ber}(p) \)
7. **else**
8. Sample \( r \sim \text{Ber}(q) \)
9. **if** \( r = 1 \) **then**
10. \( c \leftarrow c + 1 \)
11. Output \( r \)

We call this mechanism 0-favored, because for each query \((p, q)\), we always have \( q \leq p \), meaning that the algorithm is more likely to output 1 when its private input is \( b = 0 \).

We claim that Algorithm 2 with privacy parameter \( 2\varepsilon \) can simulate \( IT(\mathcal{B} : \mathcal{A}^p(S)) \) and \( IT(\mathcal{B} : \mathcal{A}^{p/\varepsilon}(S')) \). To see this, suppose that \( \mathcal{B} \) knows that he is interacting with either \( \mathcal{A}^p(S) \) or \( \mathcal{A}^{p/\varepsilon}(S') \). When \( \mathcal{B} \) prepares a hypothesis \( H \), he can compute
\[ p_0 = \Pr[\mathcal{A}^p(S) \text{ on } Test(H) \text{ outputs } \top] \]
and
\[ p_1 = \Pr[\mathcal{A}^{p/\varepsilon}(S') \text{ on } Test(H) \text{ outputs } \top]. \]

It follows that \( p_0 = p \cdot \Pr[H(S) = \top] \geq \frac{e^{-2\varepsilon}}{2} \Pr[H(S') = \top] = p_1 \), \( p_0 \in (e^{-2\varepsilon} p_1, e^{2\varepsilon} p_1) \) and \( (1 - p_0) \in (e^{-2\varepsilon} (1 - p_1), e^{2\varepsilon} (1 - p_1)) \). Therefore, sending a query \( H \) to Algorithm 1 is equivalent to sending a query \((p_0, p_1)\) to Algorithm 2.

---

6 We use \( S, S' \) to denote data sets in this subsection, since the letter “D” is reserved for divergence.
2.2.3 Analysis of the Coin Game

Algorithm 2 appears not to be differentially private. Nevertheless, it satisfies a “one-sided” stability property, which can be captured by max-divergence and Rényi divergence, as shown in the following two lemmas.

Lemma 15. Consider Algorithm 2. Fix \( B \) to be an arbitrary adversary interacting with Algorithm 2. Let \( P, Q \) denote the distributions of the interaction between \( A \) and Algorithm 2 when the private input is 0 or 1, respectively. Then, for any \( \alpha \in (1, \infty) \), we have

\[
D_\infty(P \parallel Q) \leq k \varepsilon.
\]

Lemma 16. Consider the same setup as in Lemma 15. For any \( \alpha \in (1, \infty) \), we have

\[
D_\alpha(P \parallel Q) \leq 3k \varepsilon^2.
\]

Intuitively, the two lemmas hold because for each query \((p, q)\) with \( q \leq p \), the algorithm is always more likely to output 0 when the private input is \( b = 1 \). Hence, when we consider \( D_\infty(P \parallel Q) \) (i.e., when we consider the divergence of \( P \) from \( Q \)), we can “pass” the “0” outputs “for free”. Essentially, the divergence increases only when we observe “1” outputs. The counter \( c \) counts the number of “1” outputs that we have seen so far. Since we halt the algorithm once \( c \) reaches a pre-defined threshold \( k \), naturally, one would expect that the divergence bound of \( P \) from \( Q \) grows only with \( k \) (and is independent of the number of queries). Lemmas 15 and 16 confirm this intuition.

The formal proof for Lemma 16, however, is quite delicate and involved. We defer proofs of both lemmas to the full version of the paper.

2.2.4 Wrap-up

Now, let \( B \) be an adversary interacting with \( A^p(S) \) or \( A^{p/\varepsilon}(S') \). Lemma 15 shows that for every collection of outputs \( E \), it holds that

\[
\Pr[IT(B : A^p(S)) \in E] \leq e^{2\varepsilon} \Pr[IT(B : A^{p/\varepsilon}(S')) \in E].
\]

By Hölder’s inequality and Lemma 16 (see also the conversion from Rényi DP to approximate DP, e.g., [27]), there is an absolute constant \( D \geq 1 \) such that

\[
\Pr[IT(B : A^p(S)) \in E] \leq e^{D \sqrt{\varepsilon \log(1/\delta) \varepsilon}} \Pr[IT(B : A^{p/\varepsilon}(S')) \in E] + \delta.
\]

To finish the proof of Theorem 2, we integrate over \( p \in [0, 1] \). This step is the same as the proof of Theorem 1 and we do not repeat it here.

3 Proofs for Selected Applications

In this section, we prove several selected theorems listed in Sections 1.3 and 1.4, to further illustrate the power of the private selection framework.

Notation

For \( \varepsilon, \delta > 0 \), we define the truncated Laplace mechanism \( TLap(\varepsilon, \delta) \) as follows. The support of \( TLap(\varepsilon, \delta) \) is \([-\log(1/\delta)/\varepsilon, \log(1/\delta)/\varepsilon]\]. For every \( v \in [-\log(1/\delta)/\varepsilon, \log(1/\delta)/\varepsilon] \), we have \( \Pr[TLap(\varepsilon, \delta) = v] \propto e^{-|v| \varepsilon} \). The truncated Laplace mechanism can answer a sensitivity-1 query \( f : \mathcal{X}^n \rightarrow \mathbb{R} \) by publishing \( f(D) + TLap(\varepsilon, \delta) \). It is well known that this mechanism is \((\varepsilon, \delta)\)-DP [16].
3.1 Better-Than-Median Selection

In this section, we prove Theorem 5, restated below.

Reminder of Theorem 5. For every $\alpha \in (0, \infty)$ and $\beta \in (0, 1)$, for $(\varepsilon, \delta)$-DP $A$ with $\varepsilon \in (0, 1)$ and $\delta \geq 0$, there is a better-than-median algorithm $A^*$ with confidence $1 - \beta$ that satisfies the following:

- With probability 1, $A^*$ makes at most $T$ oracle calls to $A$, where
  \[ T = \begin{cases} \left\lceil \frac{2}{\varepsilon^2} \right\rceil & \text{if } \alpha = 1 \\ \frac{5(\frac{2}{\varepsilon})^{1/\alpha} \log(1/\beta)}{\varepsilon} & \text{otherwise}. \end{cases} \]

- $A^*$ is $((2 + \alpha)\varepsilon, T\delta)$-differentially private.

Furthermore, if $A$ is $(\varepsilon, \delta)$-DP with $\delta > 0$, then $A^*$ is $(\varepsilon, T\delta)$-DP.

Proof. We design $A^*$ that initializes Algorithm 1 with $\gamma = \alpha$ and calls Selection with arguments $(k = 1, r = T, M_1 = A)$ where $T$ given by the theorem statement. If follows from Theorem 1 that $A^*$ is $((2 + \alpha)\varepsilon, \tau\delta)$-DP. As for the utility, we have shown the utility guarantee for $\alpha = 1$. When $\alpha \neq 1$, we argue as follows.

First, with probability at least $1 - (\beta/2)$, we have that $p \geq (2\beta)^{1/\alpha}$. We condition on this event. By our choice of $T$, we know that $A$ will be run for at least $t \geq \log(4/\beta)$ times with probability at least $1 - (\beta/4)$. We further condition on this event. Then, the probability that the $t$ calls of $A$ fail to yield a better-than-median solution is at most $1 - (\beta/4)$. Overall, the probability that $A^*$ fails to output a better-than-median solution is at most $1 - \beta$, as desired. ▶

3.2 Query Releasing

In this section, we prove Theorem 6, restated below.

Reminder of Theorem 6. There is a constant $C > 0$ such that the following is true. For every $\varepsilon \in (0, 1), \delta \in (0, 1/2)$ and $k \in \mathbb{N}$, there is an $(\varepsilon, \delta)$-DP algorithm that answers $k$ given sensitivity-1 queries $f_1, \ldots, f_k$, such that

\[ \Pr_f \left[ \|f - \bar{f}\|_\infty \leq C \sqrt{k \log 1/\delta + \log(1/\delta)}/\varepsilon \right] = 1, \]

where $\bar{f} := (\bar{f}_1, \ldots, \bar{f}_k)$ denotes the responses released by the algorithm.

To prove Theorem 6, we need the following theorem due to [18].

Theorem 17 ([18]). There is a constant $C_1 > 0$ such that the following is true. For every $\varepsilon \in (0, 1), \delta \in (0, 1/2)$ and $k \in \mathbb{N}$, there is an $(\varepsilon, \delta)$-DP algorithm that, given $k$ non-adaptive sensitivity-1 queries $f_1, \ldots, f_k$, returns a list of answers $(\bar{f}_1, \ldots, \bar{f}_k)$, such that

\[ \Pr_f \left[ \|f - \bar{f}\|_\infty \leq C_1 \sqrt{k \log 1/\delta}/\varepsilon \right] \geq \frac{1}{2}. \]

The following statement follows as a simple corollary of Theorem 17.

Corollary 18. There is a constant $C_2 > 0$ such that the following is true. For every $\varepsilon \in (0, 1), \delta \in (0, 1/2)$ and $k \in \mathbb{N}$, there is an $(\varepsilon, \delta)$-DP algorithm that, given $k$ non-adaptive sensitivity-1 queries $f_1, \ldots, f_k$, returns a list of answers $(\bar{f}_1, \ldots, \bar{f}_k)$ and a real $s \in \mathbb{R}^+$ satisfying the following.
With probability 1, it holds that \( \| f - \tilde{f} \|_\infty \leq s \).

With probability at least \( \frac{1}{2} \), it holds that
\[
s \leq C_2 \left( \frac{\sqrt{k \log 1/\delta}}{\epsilon} + \frac{\log(1/\delta)}{\epsilon} \right).
\]

Proof. Let \( \mathcal{A}_1 \) be an instantiation of the algorithm from Theorem 17 with privacy parameter set to \((\epsilon/2, \delta/2)\). Given a list of queries \( f_1, \ldots, f_k \), we first run \( \mathcal{A}_1 \) to produce a list of responses \( \tilde{f}_1, \ldots, \tilde{f}_k \). Let \( g = \| f - \tilde{f} \|_\infty \). We compute \( s = g + 2 \log(1/\delta) + TLap(\epsilon/2, \delta/2) \).

Finally, we release \( (\tilde{f}_1, \ldots, \tilde{f}_k) \) and \( s \). It follows that \( s \) is an upper bound for \( \| f - \tilde{f} \|_2 \) with probability 1. Moreover, since \( \| f - \tilde{f} \|_\infty \leq 4C_1 \sqrt{\frac{k \log 1/\delta}{\epsilon}} \) with the smallest associated quality score \( s \) happens with probability at least \( \frac{1}{2} \), it follows that \( s \leq \max(4C_1, 4) \cdot \left( \frac{\sqrt{k \log 1/\delta}}{\epsilon} + \frac{\log(1/\delta)}{\epsilon} \right) \) with probability at least \( \frac{1}{2} \). Since the truncated Laplace mechanism is \( (\epsilon/2, \delta/2) \)-DP, by the basic composition theorem, the whole algorithm is \((\epsilon, \delta)\)-DP.

Given Corollary 18, Theorem 6 follows by tuning parameters properly.

Proof of Theorem 6. We will combine Corollary 18 with Theorem 5. In particular, given the target privacy budget \((\epsilon, \delta)\), let \( \mathcal{A} \) be an instantiation of Corollary 18 with privacy parameters \((\epsilon/3, \delta^2/10)\). Then we use Theorem 5 on top of \( \mathcal{A} \) with \( \gamma = 1 \) and \( \beta = \delta/10 \), to select a vector \((\tilde{f}_1, \ldots, \tilde{f}_k)\) with the smallest associated quality score \( s \). Let \( C_2 = \max(4C_1, 4) \).

If \( s \leq C_2 \left( \frac{\sqrt{k \log 1/\delta}}{\epsilon} + \frac{\log(1/\delta)}{\epsilon} \right) \), then we simply output \((\tilde{f}_1, \ldots, \tilde{f}_k)\). Otherwise, we non-privately output the original vector \((f_1, \ldots, f_k)\). Since \( s \) is an upper bound of \( \| f - \tilde{f} \|_\infty \), the utility guarantee is evident.

It remains to verify that the algorithm is \((\epsilon, \delta)\)-DP. The vector \((\tilde{f}_i)_{i \in [k]}\) selected by Theorem 5 is \((\epsilon, \delta/5)\)-DP with respect to the private input. Moreover, with probability at least \( 1 - \delta/5 \), it holds that
\[
s \leq C_2 \left( \frac{\sqrt{k \log 1/\delta}}{\epsilon} + \frac{\log(1/\delta)}{\epsilon} \right).
\]

Therefore, we will do the non-private correction with probability at most \( \delta/5 \). Consequently, the whole algorithm \((\epsilon, \delta)\)-DP.

### 3.3 Stable Selection

In this section, we prove one result for stable selection (Theorem 9) to illustrate the idea. The proof of Theorem 10 is similar, and can be found in the full version of the paper.

#### Choosing Mechanism

Let \( \mathcal{F} \) be a family of 1-Lipschitz functions. Recall that \( \mathcal{F} \) is \( k \)-bounded, if for any two neighboring datasets \( D, D' \), it holds that \( \sum_{f \in \mathcal{F}} | f(D) - f(D') | \leq k \). We now prove Theorem 9, which is restated below.

Reminder of Theorem 9. Suppose \( \mathcal{F} \) is a \( k \)-bounded function family. Then, there is an \((\epsilon, \delta)\)-DP algorithm \( \mathcal{A} \) that receives a private input \( X \) and selects a function \( f \in \mathcal{F} \) such that, with probability at least \( 1 - \beta \),
\[
f(D) \geq \max_{f' \in \mathcal{F}} \{ f^*(D) \} - \mathcal{O} \left( \frac{1}{\epsilon} \log(k/\delta/\beta) \right).
\]
Proof. For each function \( f_i \in \mathcal{F} \), construct a mechanism \( \mathcal{M}_i \) as follows: \( \mathcal{M}_i \) receives the takes as input the dataset \( D \) and outputs a pair \( (i, f_i(D) + TLap(\varepsilon, \frac{5^2}{\beta \delta})) \).

We use the Selection function in Algorithm 1 with parameters \( \gamma = 1 \), \( \tau = 4 \) and mechanisms \( \{\mathcal{M}_i\}_{f_i \in \mathcal{F}} \), and output the index \( i \) returned from the Selection. To see the privacy, fix an arbitrary pair of adjacent datasets \( D, D' \). It follows that \( \sum_{f_i \in \mathcal{F}} |f(D) - f(D')| \leq k \). Moreover, it is easy to see that that \( i \)-th mechanism is \( (\varepsilon, \frac{\delta^2}{5^2}) \)-DP. Therefore, by Theorem 1, our mechanism is \( (3\varepsilon, \delta) \)-DP.

To see the utility, fix the dataset \( D \) and let \( i^* \) be the index of the best candidate. Then, with probability at least \( 1 - \beta \), at least one trial of \( \mathcal{M}_{i^*}(D) \) returns a pair \( (i, s) \) with \( s \geq f_{i^*}(D) \). We condition on this event. Let \( (j, s') \) be the result returned from Selection. It follows that \( f_j(D) \geq s' - \frac{5}{\varepsilon} \log(k/\delta) \geq s - \frac{5}{\varepsilon} \log(k/\delta) \geq f_{i^*}(D) - \frac{5}{\varepsilon} \log(k/\beta \delta) \), as desired. \( \blacktriangleleft \)

The advantage of non-random search

Looking into the proof of Theorems 9, we can find that it works by selecting from a large number of candidates \( (M_1, \ldots, M_m) \). Since the number of candidates is huge, we need to add truncated Laplace noises to balance between privacy and utility. Consequently, for each \( i \in [m] \), the best we can say about \( M_i \) is that it only satisfies \((\varepsilon, \delta_i)\)-DP with \( \delta_i > 0 \). Fortunately, fixing two neighboring inputs \( D, D' \), we can compare \( M_i(D) \) with \( M_i(D') \) for every \( i \in [m] \), and prove that the \( \delta_i \) parameters add up to at most \( O(k) \). Since our private selection considers each mechanism separately, we can bound the overall probability of privacy failure (i.e., \( \delta \)) easily. If one were to use the Liu-Talwar framework to solve the stable selection problem, then one needs to bound the privacy failure probability very carefully (perhaps via an advanced concentration inequality).

References

1. Raef Bassily, Kobbi Nissim, Adam D. Smith, Thomas Steinke, Uri Stemmer, and Jonathan R. Ullman. Algorithmic stability for adaptive data analysis. In Daniel Wichs and Yishay Mansour, editors, Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016, pages 1046-1059. ACM, 2016. doi:10.1145/2897518.2897566.
2. Amos Beimel, Kobbi Nissim, and Uri Stemmer. Private learning and sanitization: Pure vs. approximate differential privacy. Theory Comput., 12(1):1-61, 2016. doi:10.4086/toc.2016.v012a001.
3. Mark Bun, Cynthia Dwork, Guy N. Rothblum, and Thomas Steinke. Composable and versatile privacy via truncated CDP. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018, pages 74-86. ACM, 2018. doi:10.1145/3188745.3188946.
4. Mark Bun, Gautam Kamath, Thomas Steinke, and Zhiwei Steven Wu. Private hypothesis selection. IEEE Trans. Inf. Theory, 67(3):1981-2000, 2021. doi:10.1109/TIT.2021.3049802.
5. Mark Bun, Jonathan Ullman, and Salil Vadhan. Fingerprinting codes and the price of approximate differential privacy. In Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing, STOC ’14, pages 1–10, New York, NY, USA, 2014. Association for Computing Machinery. doi:10.1145/2591796.2591877.
6. Edith Cohen, Xin Lyu, Jelani Nelson, Tamás Sarlós, and Uri Stemmer. Generalized private selection and testing with high confidence, 2022. arXiv:2211.12063.
Yuval Dagan and Gil Kur. A bounded-noise mechanism for differential privacy. *CoRR*, abs/2012.03817, 2020. arXiv:2012.03817.

Jinshuo Dong, David Durfee, and Ryan Rogers. Optimal differential privacy composition for exponential mechanisms. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 2597–2606. PMLR, 13–18 July 2020. URL: https://proceedings.mlr.press/v119/dong20a.html.

David Durfee and Ryan M. Rogers. Practical differentially private top-k selection with pay-what-you-get composition. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d’Alché-Buc, Emily B. Fox, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 3527–3537, 2019. URL: https://proceedings.neurips.cc/paper/2019/hash/b139e104214a08ae3f2ebccce149c6f6e-Abstract.html.

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*, STOC ’15, pages 117–126, 2015.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. Calibrating noise to sensitivity in private data analysis. *J. Priv. Confidentiality*, 7(3):17–51, 2016.

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*, STOC ’09, pages 381–390, New York, NY, USA, 2009. Association for Computing Machinery. doi: 10.1145/1536414.1536467.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014. doi:10.1561/0400000042.

Cynthia Dwork, Guy N. Rothblum, and Salil P. Vadhan. Boosting and differential privacy. In *51th Annual IEEE Symposium on Foundations of Computer Science*, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA, pages 51–60. IEEE Computer Society, 2010. doi: 10.1109/FOCS.2010.12.

Quan Geng, Wei Ding, Ruiqi Guo, and Sanjiv Kumar. Tight analysis of privacy and utility tradeoff in approximate differential privacy. In Silvia Chiappa and Roberto Calandra, editors, *The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020, 26-28 August 2020*, Online [Palermo, Sicily, Italy], volume 108 of *Proceedings of Machine Learning Research*, pages 89–99. PMLR, 2020. URL: http://proceedings.mlr.press/v108/geng20a.html.

Badih Ghazi, Ravi Kumar, and Pasin Manurangsi. Differentially private clustering: Tight approximation ratios. In Hugo Larochelle, Marc’Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020*, virtual, 2020. URL: https://proceedings.neurips.cc/paper/2020/hash/299dc35a747eb77177d9ce10a802da2-Abstract.html.

Badih Ghazi, Ravi Kumar, and Pasin Manurangsi. On avoiding the union bound when answering multiple differentially private queries. In Mikhail Belkin and Samory Kpotufe, editors, *Conference on Learning Theory, COLT 2021, 15-19 August 2021, Boulder, Colorado, USA*, volume 134 of *Proceedings of Machine Learning Research*, pages 2133–2146. PMLR, 2021. URL: http://proceedings.mlr.press/v134/ghazi21a.html.
Generalized Private Selection and Testing with High Confidence

19 Anupam Gupta, Katrina Ligett, Frank McSherry, Aaron Roth, and Kunal Talwar. Differentially private combinatorial optimization. In Moses Charikar, editor, *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, USA, January 17-19, 2010*, pages 1106–1125. SIAM, 2010. doi:10.1137/1.9781611973075.90.

20 Moritz Hardt and Guy N. Rothblum. A multiplicative weights mechanism for privacy-preserving data analysis. In *51st Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, pages 61–70. IEEE Computer Society, 2010. doi:10.1109/FOCS.2010.85.

21 Haim Kaplan, Katrina Ligett, Yishay Mansour, Moni Naor, and Uri Stemmer. Privately learning thresholds: Closing the exponential gap. In Jacob D. Abernethy and Shivani Agarwal, editors, *Conference on Learning Theory, COLT 2020, 9-12 July 2020, Virtual Event [Graz, Austria]*, volume 125 of *Proceedings of Machine Learning Research*, pages 2263–2285. PMLR, 2020. URL: http://proceedings.mlr.press/v125/kaplan20a.html.

22 Haim Kaplan, Yishay Mansour, and Uri Stemmer. The sparse vector technique, revisited. In Mikhail Belkin and Samory Kpotufe, editors, *Conference on Learning Theory, COLT 2021, 15-19 August 2021, Boulder, Colorado, USA*, volume 134 of *Proceedings of Machine Learning Research*, pages 2747–2776. PMLR, 2021. URL: http://proceedings.mlr.press/v134/kaplan21a.html.

23 Jingcheng Liu and Kunal Talwar. Private selection from private candidates. In Moses Charikar and Edith Cohen, editors, *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, pages 298–309. ACM, 2019. doi:10.1145/3313276.3316377.

24 Min Lyu, Dong Su, and Ninghui Li. Understanding the sparse vector technique for differential privacy. *Proc. VLDB Endow.*, 10(6):637–648, 2017. doi:10.14778/3055330.3055331.

25 Ryan McKenna and Daniel Sheldon. Permute-and-flip: A new mechanism for differentially private selection. In *Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS’20*, Red Hook, NY, USA, 2020. Curran Associates Inc.

26 Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007), October 20-23, 2007, Providence, RI, USA, Proceedings*, pages 94–103. IEEE Computer Society, 2007. doi:10.1109/FOCS.2007.41.

27 Ilya Mironov. Rényi differential privacy. In *30th IEEE Computer Security Foundations Symposium, CSF 2017, Santa Barbara, CA, USA, August 21-25, 2017*, pages 263–275. IEEE Computer Society, 2017. doi:10.1109/CSF.2017.11.

28 Jack Murtagh and Salil P. Vadhan. The complexity of computing the optimal composition of differential privacy. *Theory Comput.*, 14(1):1–35, 2018. doi:10.4086/toc.2018.v014a008.

29 Sewoong Oh and Pramod Viswanath. The composition theorem for differential privacy. *CoRR*, abs/1311.0776, 2013. arXiv:1311.0776.

30 Nicolas Papernot and Thomas Steinke. Hyperparameter tuning with Rényi differential privacy. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022. URL: https://openreview.net/forum?id=7OLB1pp9DF.

31 Gang Qiao, WeiJie J. Su, and Li Zhang. One-shot differentially private top-k selection. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 8672–8681. PMLR, 2021. URL: http://proceedings.mlr.press/v139/qiao21b.html.

32 Aaron Roth and Tim Roughgarden. Interactive privacy via the median mechanism. In Leonard J. Schulman, editor, *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 765–774. ACM, 2010. doi:10.1145/1806689.1806794.
Thomas Steinke and Jonathan Ullman. Tight lower bounds for differentially private selection. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 552–563, 2017. doi:10.1109/FOCS.2017.57.

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

Thomas Steinke and Jonathan R. Ullman. Between pure and approximate differential privacy. *J. Priv. Confidentiality*, 7(2), 2016. doi:10.29012/jpc.v7i2.648.

Salil P. Vadhan. The complexity of differential privacy. In Yehuda Lindell, editor, *Tutorials on the Foundations of Cryptography*, pages 347–450. Springer International Publishing, 2017. doi:10.1007/978-3-319-57048-8_7.

Yuqing Zhu and Yu-Xiang Wang. Improving sparse vector technique with Renyi differential privacy. In Hugo Larochelle, Marc’Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL: https://proceedings.neurips.cc/paper/2020/hash/e9bf14a419d77534105016f5ec122d62-Abstract.html.