On the Optimality of the Exponential Mechanism

Francesco Aldà and Hans Ulrich Simon

Horst Görtz Institute for IT Security and Faculty of Mathematics
Ruhr-Universität Bochum, Universitätsstraße 150, 44801 Bochum, Germany
{francesco.alda,hans.simon}@rub.de

Abstract. In this work, we investigate one of the most renowned tools used in differential privacy, namely the exponential mechanism. We first study the optimality of the error introduced by the exponential mechanism in the average-case scenario, when the input/output universe of the mechanism can be modeled as a graph where each node is associated with a database. By leveraging linear programming theory, we provide some regularity conditions on the graph structure under which the exponential mechanism minimizes the average error. Moreover, we give a toy example in which the optimality is preserved (up to a constant factor) even if these regularity conditions hold only to a certain extent. Finally, we prove the worst-case optimality of the exponential mechanism when it is used to release the output of a sorting function.

1 Introduction

Differential privacy [8] is a highly popular paradigm for privacy-preserving statistical analysis. It ensures privacy by limiting the influence of an individual input datum on the released information. In addition to the rigorous privacy guarantees provided, the recognition of this framework can be traced back to some crucial factors: the composition property which permits to combine differentially private mechanisms while controlling privacy degradation, and the existence of very simple tools which easily endorse its adoption. The Laplace [8] and the exponential [19] mechanism represent the perfect example. While the Laplace mechanism provides differential privacy to vector-valued functions, the exponential mechanism is intentionally designed for applications where the response set can be arbitrary and possibly non-numeric [19]. If we ignore the efficiency issues that this algorithm inherently has, it has proved extremely successful in a number of applications, from privately generating synthetic databases that can accurately answer a large class of queries [4], to private PAC-learning [16]. Moreover, it has been shown to outperform the accuracy guarantees provided by the Laplace mechanism in some numeric settings [3].

In this paper, we first investigate under which conditions the exponential mechanism is optimal in terms of the average-case error. We consider the setting where the input and output universe of a privacy mechanism coincide and can be modeled as a graph, where each node is associated with a database, and adjacent nodes correspond to neighboring databases. The optimal privacy mechanism can then be expressed as the solution of a linear program, where we seek to minimize the average error introduced by the mechanism subject to the constraints induced by differential privacy. We show that, if the induced graph has a transitive automorphism group and a so-called regular layer sequence, then the exponential mechanism is actually optimal, i.e., its solution coincides with that of the optimal mechanism. We then provide a toy example in which this result holds (up to a constant factor) even if the aforementioned conditions are met only to a large extent. Finally, we introduce the sorting function and show that the error introduced by the exponential mechanism is actually optimal in the worst-case. We underline that this last result carries over and extends the analysis discussed in a work currently under review [1].

Related Work. A general upper bound on the error introduced by the exponential mechanism is given by McSherry and Talwar [19]. Lower bounds in differential privacy have been extensively studied and a range of techniques for proving lower bounds have been introduced [13,18,6,12,7,1]. The optimality of differentially private mechanisms has been the subject of recent studies. Kairouz et al. [15] introduce a family of mechanisms which contains a utility-maximizer under the local model of privacy. Koufogiannis et al. [17] investigate the optimality of the Laplace mechanism under the Lipschitz privacy framework. In particular, they show that the Laplace mechanism is optimal for identity queries in terms of the mean-squared error, when privacy is guaranteed with respect to the L_1-norm. Geng et al. [10] show that the mean-squared error introduced by the staircase mechanism is optimal for low-dimensional queries. Linear programming theory can be leveraged to show lower bounds on the error needed for achieving any meaningful privacy guarantee [9,6]. Hsu et al. [14] investigate how to solve a linear program under differential privacy. Hardt and Talwar [13] exploit linear programming theory to show tight upper and lower bounds on the amount of noise needed to provide differential privacy for r linear queries on databases in $\mathbb{R}^N$. Our contribution is mostly related to the work of Ghosh et al. [11] and Brenner and Nissim [5]. In their paper, Ghosh et al. [11] consider Bayesian *information consumers* that wish to compute the number of entries in a database satisfying a given predicate. An information consumer is characterized by a prior belief and a loss-function, which quantify the consumer's side knowledge and the quality of the answer. Introducing a linear program modeling the privacy constraints, they show that a discrete variant of the Laplace mechanism enables optimality (after a deterministic post-processing of the output) for *all* Bayesian information consumers. Such a mechanism is usually referred to as universally optimal. In a follow up work, Brenner and Nissim [5] show that universally optimal mechanisms for Bayesian consumers are extremely rare, proving that they essentially exist only for a single count query. Their proof makes use of a so-called *privacy constraint graph*, where the vertices correspond to the values of the output space, and the edges correspond to pairs of values resulting by applying the query function to neighboring databases. In contrast to [11] and [5], we restrict our attention to a single information consumer who has a uniform prior over the input/output space and measures the loss in terms of the record-exchange metric. We then study under which conditions on the structure of the privacy constraint graph the solution of the optimal differentially private mechanism (modeled by a linear program similar to the one introduced by Ghosh et al. [11]) coincides with the solution that the exponential mechanism delivers.

2 Preliminaries

Let $\mathcal{X}$ be a domain. A database $\mathcal{D}$ is an N-dimensional vector over $\mathcal{X}$, i.e. $\mathcal{D} \in \mathcal{X}^N$. N is referred to as the size of the database $\mathcal{D}$. Two databases $\mathcal{D}, \mathcal{D}'$ are said to be *neighboring*, denoted $\mathcal{D} \sim \mathcal{D}'$, if they can be obtained from each other by a single record exchange.

Definition 1 ([8]). *Let $\mathcal{X}$ be a domain and $\mathcal{R}$ be a (possibly infinite) set of responses. A random mechanism $\mathcal{M} : \mathcal{X}^N \to \mathcal{R}$ is said to provide* ε-differential privacy *for $\varepsilon > 0$ if, for every pair $(\mathcal{D}, \mathcal{D}')$ of neighboring databases and for every measurable $S \subseteq \mathcal{R}$, we have*

$$\Pr[\mathcal{M}(\mathcal{D}) \in S] \leq e^{\varepsilon} \cdot \Pr[\mathcal{M}(\mathcal{D}') \in S] \ .$$

The exponential mechanism [19] is a well-known tool for achieving differential privacy. Let $u \colon \mathcal{X}^N \times \mathcal{R} \to \mathbb{R}$ be a *utility* function, mapping a database/output pair to a score. Given a database $\mathcal{D} \in \mathcal{X}^N$, the exponential mechanism defines a probability distribution over $\mathcal{R}$ weighted according to the utility function $u(\mathcal{D}, \cdot)$.

Definition 2 ([19]). *Let $u \colon \mathcal{X}^N \times \mathcal{R} \to \mathbb{R}$ and $\varepsilon > 0$. The exponential mechanism $\mathcal{M}_{exp} \colon \mathcal{X}^N \to \mathcal{R}$ assigns to $s \in \mathcal{R}$ a probability density proportional to*

$$\exp\left(\frac{\varepsilon \cdot u(\mathcal{D}, s)}{S(u)}\right) \ , \tag{1}$$

where $S(u) = \sup_{s\in\mathcal{R}} \sup_{\mathcal{D}\sim\mathcal{D}'} |u(\mathcal{D}, s) - u(\mathcal{D}', s)|$ is the sensitivity *of u. It then returns a value sampled from such distribution.*

We briefly note that the definition in [19] is slightly more general and has an additional factor $\mu(s)$ in (1), which represents a prior distribution on $\mathcal{R}$. In this paper, we deal with a uniform prior and have therefore omitted $\mu(s)$ from Definition 2.

Lemma 1 ([19]). *The exponential mechanism provides 2ε-differential privacy.*

In several cases (see for example the unit demand auction setting in [19]) the factor 2 in the statement of Lemma 1 can be removed, strengthening the privacy guarantees to ε-differential privacy.

3 Optimal Mechanisms and Linear Programming

Let $G = (\mathcal{K}, E)$ denote a graph with $K = |\mathcal{K}|$ nodes and diameter D. Intuitively, we should think of each node $x \in \mathcal{K}$ as a piece of information associated with a database $\mathcal{D}$. Moreover, we should think of adjacent nodes in G as nodes whose underlying databases are neighbored in the sense that they can be obtained from each other by a single record exchange. Hence a node y has distance d from another node x iff d is the smallest number of record exchanges which transforms the database underlying y into the database underlying x. We consider the following special situation:

- The (randomized) mechanisms $\mathcal{M}$ under investigation should provide ε-differential privacy and, given a node $x \in \mathcal{K}$, they should return another node in $\mathcal{K}$ (the choice of which depends on $\mathcal{M}$'s internal randomization).
- The cost (= negated utility) of an output y, given input x, is defined as the distance between x and y in G, which is denoted as $d(x, y)$. We will refer to this distance measure as the *record-exchange metric*. Note that $|d(x, y) - d(x', y)| \leq 1$ holds for all $x, x', y \in \mathcal{K}$ such that x and x' (resp. their underlying databases) are neighbored. Thus $-d$ (viewed as a utility function) has sensitivity 1.

Note that the record-exchange metric coincides with what is called "geodesic distance" w.r.t the graph G in some papers.

We consider two examples where, in both cases, the record-exchange metric coincides with 1/2 times the L_1-metric.

Example 1. Suppose that the nodes in G represent histograms with N users and T types of records (briefly called (N, T)-histograms hereafter), i.e., we may identify a node $x \in \mathcal{K}$ with a vector $(v_1, \ldots, v_T)$ such that $N = \sum_{t=1}^{T} v_t$ and v_t is the number of users whose record is of type t. Note that the record-exchange metric satisfies $d(x, y) = \frac{1}{2}\|y - x\|_1$ because each record-exchange can decrease the L_1-distance between two histograms by an amount of 2 (but not more).

Example 2. Suppose that the nodes in G represent sorted (N, T)-histograms, i.e., we may identify a node $x \in \mathcal{K}$ with a sorted sequence $v_1 \geq \ldots \geq v_T$ such that $\sum_{t=1}^{T} v_t = N$. Here v_1 (resp. v_2 and so on) denotes the number of users whose record occurs most often (resp. 2nd most often and so on) in the database. Alternatively, we may be interested in the $r \leq T$ largest values of $v_1 \geq \ldots \geq v_T$ only, i.e., we identify a node $x \in \mathcal{K}$ with the initial segment $v_1 \geq \ldots, \geq v_r$ of the full sequence $v_1 \geq \ldots \geq v_T$.

In this section, we investigate under which conditions the exponential mechanism is optimal in the sense of incurring the smallest possible expected error (measured in terms of the record-exchange metric) where expectation is taken over the (uniformly distributed) inputs $x \in_R \mathcal{K}$ and over the internal randomization of the mechanism. We start by introducing several linear programs. The optimal solution of the first linear program we consider, denoted LP[1] below, corresponds to the solution of the optimal ε-differentially private mechanism. Another linear program, denoted LP[3] below, has an optimal solution which coincides with the one given by the exponential mechanism. We then provide some regularity conditions on the graph G under which an optimal solution of LP[3] also optimizes LP[1] (so that the exponential mechanism is optimal whenever the regularity conditions are valid).

We can now continue with our general discussion. Note that a (randomized) mechanism $\mathcal{M}$ with inputs and outputs taken from $\mathcal{K}$ is formally given by probability parameters $p(y|x)$ denoting the probability of returning $y \in \mathcal{K}$ when given $x \in \mathcal{K}$ as input. Since, for each x, $p(y|x)$ is a distribution on $\mathcal{K}$, we have

$$(\forall x, y \in \mathcal{K} :\ p(y|x) \geq 0) \wedge \left(\forall x \in \mathcal{K} :\ \sum_{y \in \mathcal{K}} p(y|x) = 1\right) \ . \tag{2}$$

Moreover, if $\mathcal{M}$ provides ε-differential privacy, we have

$$\forall y \in \mathcal{K}, \forall \{x, x'\} \in E :\ p(y|x') \geq e^{-\varepsilon} \cdot p(y|x) \ . \tag{3}$$

Conversely, every choice of these probability parameters that satisfies (2) and (3) represents a mechanism that provides ε-differential privacy.

Suppose that $\mathcal{M}$ is given by its probability parameters $p = (p(y|x))$ as described above. The average distance between $x \in \mathcal{K}$ and the output $y \in \mathcal{K}$, returned by $\mathcal{M}$ when given x as input, is then given as follows:

$$f^G(p) = \frac{1}{K} \cdot \sum_{x \in \mathcal{K}} \sum_{y \in \mathcal{K}} p(y|x) d(x, y) \ . \tag{4}$$

Let $S_d = S_d(y)$ denote the set of all nodes in $\mathcal{K}$ with distance d to y (the d-th layer of G w.r.t. start node y). Then

$$f^G(p) = \frac{1}{K} \cdot \sum_{y \in \mathcal{K}} f^G_y(p) \ \text{ for } \ f^G_y(p) = \sum_{d=0}^{D} \sum_{x \in S_d(y)} p(y|x) d \ . \tag{5}$$

We pursue the goal to find an ε-differentially private mechanism $\mathcal{M}$ that minimizes $d(x, y)$ on the average. For this reason, we say that a mechanism $\mathcal{M}_*$ with probability parameters p_* is *optimal* w.r.t. G if p_* is a minimizer of $f^G(p)$ among all p that satisfy (2) and (3). It is obvious from our discussion that the probability parameters $p_*(y|x)$ representing an optimal mechanism w.r.t. G are obtained by solving the following linear program:

$$\mathrm{LP}[1] : \mathbf{min}_{p=(p(y|x))_{x,y \in \mathcal{K}}}\ f^G(p)\ \mathbf{s.t.}\ (2) \text{ and } (3) \ .$$

We will refer to this linear program as $\mathrm{LP}^G[1]$ whenever we want to stress the dependence on the underlying graph G. We now bring into play the following modifications of the condition (2):

$$(\forall x, y \in \mathcal{K} : p(y|x) \geq 0) \wedge \left(\sum_{x \in \mathcal{K}} \sum_{y \in \mathcal{K}} p(y|x) = K\right) \ . \tag{6}$$

$$(\forall x, y \in \mathcal{K} : p(y|x) \geq 0) \wedge \left(\forall y \in \mathcal{K} :\ \sum_{x \in \mathcal{K}} p(y|x) = 1\right) \ . \tag{7}$$

Note that (7) implies (6). Consider the following relatives of $\mathrm{LP}^G[1]$:

$$\mathrm{LP}[2] : \mathbf{min}_{p=(p(y|x))_{x,y\in\mathcal{K}}} f^G(p) \ \mathbf{s.t.}\ (6) \text{ and } (3)\ \ ;$$
$$\mathrm{LP}[3] : \mathbf{min}_{p=(p(y|x))_{x,y\in\mathcal{K}}} f^G(p) \ \mathbf{s.t.}\ (7) \text{ and } (3)\ \ .$$

As for $\mathrm{LP}^G[1]$, we will use the notations $\mathrm{LP}^G[2]$ and $\mathrm{LP}^G[3]$ to stress the dependence on the underlying graph G. Given a graph $G = (\mathcal{K}, E)$, a permutation σ of $\mathcal{K}$ is called *automorphism* if, for all $x, y \in \mathcal{K}$, $\{x, y\} \in E \Leftrightarrow \{\sigma(x), \sigma(y)\} \in E$. The set of all automorphisms of $\mathcal{K}$, under the operation of composition of functions, forms a group called the automorphism group of G. Such a group is called *transitive* if, for every $x, y \in \mathcal{K}$, there exists an automorphism σ of $\mathcal{K}$ such that $\sigma(x) = y$.

Lemma 2. *Suppose that the graph G has a transitive automorphism group. Then every feasible solution p for $\mathrm{LP}^G[2]$ can be transformed into another feasible solution p' such that $f^G(p') \le f^G(p)$ and p' satisfies (7).*

Proof. Let p be any feasible solution for $\mathrm{LP}^G[2]$. For every $y \in \mathcal{K}$, let $K_y(p) = \sum_{x\in\mathcal{K}} p(y|x)$. According to (6), we have $\sum_{y\in\mathcal{K}} K_y(p) = K$. Define

$$\bar{p}(y|x) = \frac{1}{K_y(p)} p(y|x) \ \text{ and } \ \bar{f}_y(p) = \sum_{d=0}^{D} \sum_{x\in S_d(y)} \bar{p}(y|x)d$$

and note that $\bar{p}$ satisfies (3) and (7). We may now write $f^G(p)$ as follows:

$$f^G(p) = \sum_{y\in\mathcal{K}} \frac{K_y(p)}{K} \cdot \bar{f}_y(p)\ \ .$$

Thus $f^G(p)$ can be interpreted as the average of the cost terms $f_y(p)$ where the term $f_y(p)$ is chosen with probability $K_y(p)/K$. According to the pigeonhole principle, there exists $y^* \in \mathcal{K}$ such that $\bar{f}_{y^*}(p) \le f^G(p)$. Our strategy is to use the automorphism of G for building a new (and superior) feasible solution p' whose components contain K duplicates of the parameter collection $(\bar{p}(y^*|x)_{x\in\mathcal{K}})$. To this end, let σ_y be the automorphism which maps y to y^* and define

$$p'(y|x) = \bar{p}(y^*|\sigma_y(x))\ \ .$$

Note that $x \in S_d(y)$ if and only if $\sigma_y(x) \in S_d(y^*)$. Obviously, $p' \ge \mathbf{0}$ and, for every $y \in \mathcal{K}$, we have

$$K_y(p') = \sum_{x\in\mathcal{K}} p'(y|x) = \sum_{x\in\mathcal{K}} \bar{p}(y^*|\sigma_y(x)) = \sum_{x\in\mathcal{K}} \bar{p}(y^*|x) = 1\ \ .$$

This shows that p' satisfies (7). Moreover, p' satisfies (3) since, for every $y \in \mathcal{K}$ and every $\{x, x'\} \in E$, we have

$$e^{-\varepsilon} \cdot p'(y|x) = e^{-\varepsilon} \cdot \bar{p}(y^*|\sigma_y(x)) \le \bar{p}(y^*|\sigma_y(x')) = p'(y|x')\ \ ,$$

where the inequality follows from the fact that $\bar{p}$ satisfies (3) and σ_y is an automorphism. The following calculation shows that $f_y(p') = \bar{f}_{y^*}(p)$ holds for every $y \in \mathcal{K}$:

$$\begin{aligned} f_y(p') &= \sum_{d=0}^{D} \sum_{x \in S_d(y)} p'(y|x)d = \sum_{d=0}^{D} \sum_{x \in S_d(y)} \bar{p}(y^*|\sigma_y(x))d \\ &= \sum_{d=0}^{D} \sum_{x \in S_d(y^*)} \bar{p}(y^*|x)d = \bar{f}_{y^*}(p) \enspace . \end{aligned}$$

We now obtain

$$f^G(p') = \frac{1}{K} \cdot \sum_{y \in \mathcal{K}} f_y(p') = \bar{f}_{y^*}(p) \le f^G(p) \enspace ,$$

which concludes the proof. □

The following result is an immediate consequence of Lemma 2.

Corollary 1. *The optimal values of the problems* LP[2] *and* LP[3] *coincide. Moreover, every optimal solution for* LP[3] *is an optimal solution for* LP[2].

We say that the graph G *has a regular layer sequence w.r.t.* $y \in \mathcal{K}$ if, for all d and for all $x, x' \in S_d(y)$, the nodes x and x' have the same number of neighbors in $S_{d-1}(y)$ and the same number of neighbors in $S_{d+1}(y)$. Let $E[y] = E \cap (S_d(y) \times S_{d+1}(y))$, i.e., $E[y]$ contains the edges in E which connect two nodes in subsequent layers (but excludes the edges which connect two nodes in the same layer).

Lemma 3. *Suppose that the graph $G = (\mathcal{K}, E)$ has a transitive automorphism group and a regular layer sequence w.r.t. any $y \in \mathcal{K}$. Then the problems* $\mathrm{LP}^G[2]$ *and* $\mathrm{LP}^G[3]$ *have an optimal solution that satisfies*

$$\forall y \in \mathcal{K}, \forall (x, x') \in E[y] : \; p(y|x') \ge e^{-\varepsilon} \cdot p(y|x) \tag{8}$$

with equality.

Proof. The problem LP[3] decomposes into $K = |\mathcal{K}|$ independent subproblems, one subproblem $\mathrm{LP}(y)$ for each fixed choice of $y \in \mathcal{K}$:

$$\begin{aligned} \mathrm{LP}(y) : \mathbf{min}_{p=(p(y|x))_{x \in \mathcal{K}}} f_y^G(p) &= \sum_{d=0}^{D} \left(\sum_{x \in S_d(y)} p(y|x) \right) d \\ \mathbf{s.t.}\ (p \ge \mathbf{0}) \wedge \left(\sum_{x \in \mathcal{K}} p(y|x) = 1 \right) &\wedge \big(\forall \{x, x'\} \in E : p(y|x') \ge e^{-\varepsilon} \cdot p(y|x) \big) \enspace . \end{aligned}$$

Let LP[5] (the numbering will become clear in Section 4) be the linear program that is obtained from $\mathrm{LP}(y)$ by substituting the weaker constraint

$$\forall (x, x') \in E[y] : p(y|x') \ge e^{-\varepsilon} \cdot p(y|x)$$

for

$$\forall \{x, x'\} \in E : p(y|x') \geq e^{-\varepsilon} \cdot p(y|x) \ .$$

In Section 4 we will prove the following result:

Claim 1. *If $G(y)$ has a regular layer sequence, then* LP[5] *has an optimal solution with the following properties:*

1. *The parameter vector $(p(y|x))_{x \in \mathcal{K}}$ (with a fixed choice of y) assigns the same probability mass to all nodes x taken from the same layer.*
2. *For every $(x, x') \in E[y]$, it satisfies the constraint $p(y|x') \geq e^{-\varepsilon} \cdot p(y|x)$ with equality.*

It immediately follows that this optimal solution is also an optimal solution for LP(y), which completes the proof. □

The proof of Claim 1 is lengthy and will therefore be given later. See Lemma 5 in Section 4. Recall that $d(x, y)$ denotes the distance between x and y w.r.t. the record-exchange metric. Here comes the main result of this section which essentially states that the exponential mechanism is optimal under the assumptions made in Lemma 3.

Theorem 1. *Under the same assumptions as in Lemma 3, the following holds. An optimal mechanism for* LP^G[1] *(and even for* LP^G[2] *and for* LP^G[3]*) is obtained by setting*

$$\forall x, y \in \mathcal{K} : \ p(y|x) \propto \exp(-\varepsilon \cdot d(x, y)) \ .$$

Proof. Let p be the optimal solution for LP^G[2] and LP^G[3] that satisfies (8) with equality so that

$$\forall y \in \mathcal{K}, \forall (x, x') \in E[y] : \ p(y|x') = e^{-\varepsilon} \cdot p(y|x) \ .$$

Unrolling this recursion, we get

$$p(y_0|x_0) = \frac{\exp(-\varepsilon \cdot d(x_0, y_0))}{\sum_{x \in \mathcal{K}} \exp(-\varepsilon \cdot d(x, y_0))} \ .$$

The transitivity of the automorphism group of G implies that

$$\forall x_0, y_0 \in \mathcal{K} : \ \sum_{x \in \mathcal{K}} \exp(-\varepsilon \cdot d(x, y_0)) = \sum_{y \in \mathcal{K}} \exp(-\varepsilon \cdot d(y, x_0)) \ .$$

It follows that $p(y_0|x_0) = p(x_0|y_0)$. As a feasible solution of LP^G[3], p satisfies (7). Since $p(y_0|x_0) = p(x_0|y_0)$, it must also satisfy (2). Thus p is a feasible solution for LP^G[1]. Since it is even optimal among the feasible solutions of the relaxation LP^G[2], it must be optimal for LP^G[1]. □

4 Proof of Claim 1 and Additional Remarks on LP[5]

Recall that $G = (\mathcal{K}, E)$ denotes a graph with $K = |\mathcal{K}|$ nodes and diameter D. Fix some $y \in \mathcal{K}$ and call it the "start node". Recall that $S_d = S_d(y)$ is the set of all nodes in $\mathcal{K}$ with distance d to y (the d-th layer in G). The cardinality of $S_d(y)$ is denoted by $s_d(y)$, or simply by s_d. For instance, $S_0 = \{y\}$ and S_1 is the set of all neighbors of y in G. An edge $e \in E$ either connects two nodes in subsequent layers or it connects two nodes in the same layer. Let again $E[y] \subseteq E$ be the set of edges of the former kind and let $G[y] = (\mathcal{K}, E[y])$. In other words, $G[y]$ is the layered graph that contains all shortest paths to the start node y. We consider an edge in $E[y]$ as being directed away from y, i.e., $(x, x') \in E[y]$ implies that $x \in S_d$ and $x' \in S_{d+1}$ for some $d \in [0 : D-1]$. Note that $E[y]$ naturally partitions into the (disjoint) union of $E_0, E_1, \ldots, E_{D-1}$ where $E_d = E[y] \cap (S_d \times S_{d+1})$. Let $0 < \gamma < 1$ denote a constant scaling factor. In this section, we consider the following two linear optimization problems:

Linear Program 4 (LP[4])		Linear Program 5 (LP[5])	
min$_{p=(p_d)}$	$f_4(p) = \sum_{d=0}^{D} s_d p_d d$	**min**$_{p=(p_x)}$	$f_5(p) = \sum_{d=0}^{D} \left(\sum_{x \in S_d} p_x \right) d$
s.t.	$p \geq \mathbf{0}$, $\sum_{d=0}^{D} s_d p_d = 1$,	**s.t.**	$p \geq \mathbf{0}$, $\sum_{x \in \mathcal{K}} p_x = 1$,
(C4)	$\forall d \in [0 : d-1] : p_{d+1} \geq \gamma \cdot p_d$.	**(C5)**	$\forall (x, x') \in E[y] : p_{x'} \geq \gamma \cdot p_x$.

In other words, we would like to find a probability distribution on $\mathcal{K}$ that minimizes the average distance to the start node y subject to (C4) resp. (C5). In Problem LP[5], we can assign individual probabilities to all nodes whereas, in Problem LP[4], we have to assign the same probability p_d to all nodes in the d-th layer S_d (so that the total probability mass assigned to S_d equals $s_d p_d$). Note that LP[5] yields the problem that occurs under the same name in the proof of Lemma 3 provided that we set $\gamma = e^{-\varepsilon}$ and $p_x = p(y|x)$.

As for LP[4], it is intuitively clear that we should move as much probability mass as possible to layers close to the start node y. Thus the following result (whose proof is omitted) does not come as surprise:

Lemma 4. LP[4] *is bounded and feasible. Moreover, there is a unique optimal solution that satisfies all constraints in (C4) with equality.*

Recall that G with start node y is said to have a regular layer sequence if nodes in the same layer of $G[y]$ have the same in-degree and the same out-degree. The next result is essentially a reformulation of Claim 1 from Section 3.

Lemma 5. LP[5] *is bounded and feasible. Moreover, if* $G[y] = (\mathcal{K}, E[y])$ *has a regular layer sequence, then* LP[5] *has an optimal solution that, first, assigns the same probability mass to all nodes in the same layer, and, second, satisfies all constraints in (C5) with equality.*

Proof. Clearly LP[5] is bounded. Showing the existence of a feasible solution is straightforward and hence omitted. Thus we have only to show that LP[5] has an optimal solution that satisfies all constraints in (C5) with equality. Call a feasible solution $p = (p_x)$ of LP[5] *normalized* if p assigns the same probability mass to all nodes in the same layer, say $p_x = \bar{p}_d$ for every node x in layer d. As for normalized feasible solutions, LP[5] collapses to LP[4]. According to Lemma 4, there is a unique optimal solution among all normalized feasible solutions of LP[5] that satisfies all constraints in (C5) with equality.[1] Thus, we now have to show that every feasible solution can be normalized without increasing its cost. To this end, let $p = (p_x)$ denote a fixed but arbitrary feasible solution for LP[5]. For $d = 0, 1, \ldots, D$, we set $\bar{p}_d = \frac{1}{s_d} \sum_{x \in S_d} p_x$, i.e., $\bar{p}_d$ is the probability mass assigned by p to nodes in S_d on the average. We claim that setting $p'_x = \bar{p}_d$ for every node $x \in S_d$ yields a normalized feasible solution of the same cost as p. Clearly $p' \geq \mathbf{0}$. Moreover $\sum_{x \in \mathcal{K}} p'_x = \sum_{x \in \mathcal{K}} p_x = 1$ because $p \mapsto p'$ leaves the total probability mass assigned to any layer S_d unchanged. For the same reason the cost of p' coincides with the cost of p, i.e., $f_5(p') = f_5(p)$. It remains to show that p' satisfies (C5). To this end, pick any $d \in [0 : D-1]$ and any $(x, x') \in E_d$. Let $t_d^{\rightarrow}$ denote the out-degree of x (or of any other node from S_d) and let $t_{d+1}^{\leftarrow}$ denote the in-degree of x' (or of any other node from S_{d+1}). A simple double counting argument shows that

$$s_d t_d^{\rightarrow} = |E_d| = s_{d+1} t_{d+1}^{\leftarrow} \ . \tag{9}$$

The following calculation shows that $p'_{x'} \geq \gamma p'_x$:

$$\begin{aligned} p'_{x'} &= \frac{1}{s_{d+1}} \cdot \sum_{v \in S_{d+1}} p_v \\ &\overset{*}{=} \frac{1}{s_{d+1} t_{d+1}^{\leftarrow}} \cdot \sum_{v \in S_{d+1}} \sum_{u : (u,v) \in E_d} p_v \\ &\overset{(9)}{=} \frac{1}{s_d t_d^{\rightarrow}} \cdot \sum_{u \in S_d} \sum_{v : (u,v) \in E_d} p_v \\ &\geq \gamma \cdot \frac{1}{s_d t_d^{\rightarrow}} \cdot \sum_{u \in S_d} \sum_{v : (u,v) \in E_d} p_u \\ &\overset{*}{=} \gamma \cdot \frac{1}{s_d} \cdot \sum_{u \in S_d} p_u = \gamma \cdot p'_x \end{aligned}$$

The equations marked "$*$" make use of our assumption that $G[y]$ has a regular layer sequence. The whole discussion can be summarized by saying that p' is a normalized feasible solution for LP[5] and its cost equals the cost of the feasible solution p that we started with. This concludes the proof. □

Let LP[4]$^\infty$ and LP[5]$^\infty$ denote the optimization problems that result from LP[4] and LP[5], respectively, when the underlying graph $G = (\mathcal{K}, E)$ has infinitely many nodes so that the layered graph $G[y] = (\mathcal{K}, E[y])$ might have

[1] (C5) collapses to (C4) for normalized feasible solutions.

infinitely many layers $S_0, S_1, S_2, \ldots$. In the formal definition of LP[4] and LP[5], we only have to substitute ∞ for D. An inspection of the proofs of Lemmas 4 and 5 reveals that they hold, mutatis mutandis, for the problems LP[4]$^\infty$ and LP[5]$^\infty$ as well:

Corollary 2. LP[4]$^\infty$ *and* LP[5]$^\infty$ *are bounded and feasible. Moreover, there is a unique optimal solution for* LP[4]$^\infty$ *that satisfies all constraints in (C4) with equality and, if $G[y] = (\mathcal{K}, E[y])$ has a regular layer sequence, then* LP[5]$^\infty$ *has an optimal solution that satisfies all constraints in (C5) with equality.*

Example 3. Let G_1 be an infinite path $y_0, y_1, y_2, \ldots$ with start node y_0. It follows from Corollary 2 that LP[5] has an optimal solution that satisfies all constraints in (C5) with equality. This leads to the following average distance from y_0:

$$\frac{\sum_{d \geq 1} \gamma^d d}{\sum_{d \geq 0} \gamma^d} = \frac{\frac{\gamma}{(1-\gamma)^2}}{\frac{1}{1-\gamma}} = \frac{\gamma}{1-\gamma} \quad .$$

Let G_2 be the graph consisting of two infinite paths, $y_0, y_{-1}, \ldots$ and $y_0, y_1, \ldots$ both of which are starting from the start node y_0. Again Corollary 2 applies and the optimal average distance from y_0 is calculated as follows:

$$\frac{2 \cdot \sum_{d \geq 1} \gamma^d d}{1 + 2 \cdot \sum_{d \geq 1} \gamma^d} = \frac{\frac{2\gamma}{(1-\gamma)^2}}{1 + \frac{2\gamma}{1-\gamma}} = \frac{2\gamma}{1-\gamma^2} \quad . \tag{10}$$

As for finite paths, we have the following result:

Lemma 6. *Let P_ℓ be a path of length 2ℓ and let y_0 be the start node located in the middle of P_ℓ. Let $f(\ell)$ denote the optimal value that can be achieved in the linear program* LP[5] *w.r.t. to $G = P_\ell$. Then the following holds:*

1. LP[5] *has an optimal solution that satisfies all constraints in (C5) with equality so that*
$$f(\ell) = \frac{2 \cdot \sum_{d=1}^{\ell} \gamma^d d}{1 + 2 \cdot \sum_{d=1}^{\ell} \gamma^d} \quad . \tag{11}$$
2. *The function $f(\ell)$ is strictly increasing with ℓ.*
3. *We have*
$$f(\ell) > \frac{2\gamma}{1-\gamma^2} \cdot \left(1 - \gamma^\ell - \ell\gamma^\ell(1-\gamma)\right) \quad . \tag{12}$$
Moreover, if $\ell \geq \frac{s}{1-\gamma}$, then
$$f(\ell) > \frac{2\gamma}{1-\gamma^2} \cdot \left(1 - (s+1)e^{-s}\right) \quad . \tag{13}$$
4. $\lim_{\ell \to \infty} f(\ell) = \frac{2\gamma}{1-\gamma^2}$.

Proof. Let $P_\ell = y_{-\ell}, \ldots, y_{-1}, y_0, y_1, \ldots, y_\ell$.

1. Lemma 5 applies because $P_\ell[y_0]$ has a regular layer sequence.
2. An optimal solution for $P_{\ell+1}$ can be transformed into a feasible solution for P_ℓ by transferring the probability mass of the nodes $y_{-(\ell+1)}, y_{\ell+1}$ to the nodes $y_{-\ell}, y_\ell$, respectively. This transfer strictly reduces the cost. The optimal cost $f(\ell)$ that can be achieved on P_ℓ is, in turn, smaller than the cost of this feasible solution.
3. We start with the following calculation:

$$\begin{aligned}\sum_{d=1}^{\ell} \gamma^{d-1} d &= \sum_{d\geq 1} \gamma^{d-1} d - \sum_{d\geq \ell+1} \gamma^{d-1} d \\ &= \frac{1}{(1-\gamma)^2} - \gamma^\ell \cdot \sum_{d\geq 1} \gamma^{d-1}(d+\ell) \\ &= \frac{1}{(1-\gamma)^2} - \gamma^\ell \cdot \left(\frac{1}{(1-\gamma)^2} + \frac{\ell}{1-\gamma}\right) \\ &= \frac{1}{(1-\gamma)^2} \cdot \big(1 - \gamma^\ell - \ell\gamma^\ell(1-\gamma)\big)\end{aligned}$$

   Setting $F = 1 - \gamma^\ell - \ell\gamma^\ell(1-\gamma)$, it follows that

$$f(\ell) = \frac{\frac{2\gamma}{(1-\gamma)^2}}{1 + 2\cdot \sum_{d=1}^{\ell} \gamma^d} \cdot F > \frac{\frac{2\gamma}{(1-\gamma)^2}}{1 + 2\cdot \sum_{d\geq 1} \gamma^d} \cdot F \ \ .$$

   Since the latter expression differs from (10) by the factor F only, we obtain (12).
   The function $s \mapsto (s+1)e^{-s}$ is strictly monotonically decreasing for all $s \geq 0$. It suffices therefore to verify the bound (13) for $s = (1-\gamma)\ell$ so that

$$\gamma^\ell + \ell\gamma^\ell(1-\gamma) = \gamma^\ell(s+1) \ \ .$$

   Noting that

$$\gamma^\ell = \gamma^{s/(1-\gamma)} = (1-(1-\gamma))^{s/(1-\gamma)} < e^{-s} \ \ ,$$

   we may conclude that $\gamma^\ell + \ell\gamma^\ell(1-\gamma) < (s+1)e^{-s}$. From this, in combination with (12), the bound (13) is immediate.
4. The fourth assertion of Lemma 6 is immediate from the third one.

□

Even though the regularity conditions for G (transitive automorphism group and regular layer sequence) are satisfied in simple settings (for instance, when each node in the graph corresponds to a binary database of size N), we do not expect this to be the case in most applications. For example, the regularity conditions are not fully satisfied by the graph representing sorted (N, T)-histograms introduced in Example 2. However, we conjecture, first, that these conditions are approximately satisfied for very large databases and, second, that the exponential mechanism is still approximately optimal when these conditions hold approximately. At the time being, we are not able to verify this conjecture for graphs G of practical interest. In the next section, we will illustrate the kind of arguments that we plan to bring into play by presenting a very precise analysis for the simple case where the graph G actually is a long but finite path. Developing these arguments further so as to analyze more reasonable classes of graphs (e.g., graphs representing the neighborhood relation for sorted histograms) remains a subject of future research.

5 A Toy Example: the Path Graph

Throughout this section, we consider the graph $G = (\mathcal{K}, E)$ whose nodes $y_1, \ldots, y_K$ form a path of length $K - 1$. Note that G does not satisfy the regularity condition: neither has G a transitive automorphism group nor has $G[y]$ a regular layer sequence (except for y being chosen as one of the endpoints and, if K is odd, for y being chosen as the point in the middle of the path). Let $\mathrm{OPT}^G[1]$ denote the smallest cost of a feasible solution for $\mathrm{LP}^G[1]$. We will show in this section that, despite the violation of the regularity condition, the exponential mechanism comes close to optimality provided that K is "sufficiently large". The main idea for proving this is as follows. We will split the set of nodes into a "central part" (nodes separated away from the endpoints of the path) and a "peripheral part" (nodes located close to the endpoints). Then we make use of the fact that all ε-differentially private mechanisms are on the horns of the following dilemma:

- If a feasible solution $p = (p(y|x))_{x,y\in\mathcal{K}}$ puts much probability mass on peripheral nodes y, then the cost contribution of the terms $p(y|x)$ with y "peripheral" and x "central" will be large.
- If not, then the cost contribution of the terms $p(y|x)$ with y "central" will be large. The proof of this statement will exploit the fact that, if y has distance at least ℓ to both endpoints of the path, then $G[y]$ contains the path P_ℓ from Lemma 6 (with y located in the middle of P_ℓ) as a subgraph. It is then easy to argue that the term $f(\ell)$ from Lemma 6 serves as a lower bound on the cost achieved by p.

We will now formalize these ideas. Let $\ell \geq 1$ be arbitrary but fixed. We assume that $K \geq 4\ell$. We define the following sets of "peripheral" nodes:

$$\mathcal{K}_1 = \{y_1, \ldots, y_\ell\} \cup \{y_{K-\ell+1}, \ldots, y_K\} \text{ and}$$
$$\mathcal{K}_2 = \{y_1, \ldots, y_{2\ell}\} \cup \{y_{K-2\ell+1}, \ldots, y_K\} .$$

In other words, $\mathcal{K}_1$ (resp. $\mathcal{K}_2$) contains all nodes that have a distance of at most $\ell - 1$ (resp. $2\ell - 1$) to one of the endpoints y_1 and y_K. The complements of these sets are denoted $\bar{\mathcal{K}}_1$ and $\bar{\mathcal{K}}_2$, respectively. Note that each node in $\bar{\mathcal{K}}_1$ (resp. $\bar{\mathcal{K}}_2$) has a distance of at least ℓ (resp. 2ℓ) to both of the endpoints. Moreover, any point in $\mathcal{K}_1$ has distance of at least ℓ to any node in $\bar{\mathcal{K}}_2$. For every set $M \subseteq \mathcal{K} \times \mathcal{K}$, we define

$$P(M) = \sum_{(x,y)\in M} p(x,y) = \frac{1}{K} \cdot \sum_{(x,y)\in M} p(y|x) ,$$

i.e., $P(M)$ is the total probability mass assigned to pairs $(x,y) \in M$ if $x \in \mathcal{K}$ is uniformly distributed and y has probability $p(y|x)$ conditioned to x. Then $P(\mathcal{K} \times \mathcal{K}_1)$ denotes the total probability assigned to pairs (x,y) with $y \in \mathcal{K}_1$. The total mass of pairs from $\bar{\mathcal{K}}_2 \times \mathcal{K}_1$ can then be bounded from below as follows:

$$\begin{aligned} P(\bar{\mathcal{K}}_2 \times \mathcal{K}_1) &= P(\bar{\mathcal{K}}_2 \times \mathcal{K}) - P(\bar{\mathcal{K}}_2 \times \bar{\mathcal{K}}_1) \geq P(\bar{\mathcal{K}}_2 \times \mathcal{K}) - P(\mathcal{K} \times \bar{\mathcal{K}}_1) \\ &= \left(1 - \frac{4\ell}{K}\right) - (1 - P(\mathcal{K} \times \mathcal{K}_1) = P(\mathcal{K} \times \mathcal{K}_1) - \frac{4\ell}{K} \ . \end{aligned}$$

Since $p(x,y) = p(y|x)/K$, we may rewrite the cost function $f^G(p)$ from (4) as follows:

$$f^G(p) = \sum_{(x,y) \in \mathcal{K} \times \mathcal{K}} p(x,y) d(x,y) \ .$$

Since, as mentioned above already, $d(x,y) \geq \ell$ holds for all pairs $(x,y) \in \bar{\mathcal{K}}_2 \times \mathcal{K}_1$, we obtain a first lower bound on $f^G(p)$:

$$f^G(p) \geq P(\bar{\mathcal{K}}_2 \times \mathcal{K}_1) \cdot \ell \geq \left(P(\mathcal{K} \times \mathcal{K}_1) - \frac{4\ell}{K}\right) \cdot \ell \ . \tag{14}$$

The lower bound (14) is induced by the elements y taken from the "peripheral region" $\mathcal{K}_1$. In the next step, we derive a lower bound that is induced by the elements y taken from the "central region" $\bar{\mathcal{K}}_1$. We remind the reader of the short notation

$$K_y(p) = \sum_{x \in \mathcal{K}} p(y|x) \quad \text{and} \quad \bar{f}_y(p) = \sum_{x \in \mathcal{K}} \frac{p(y|x)}{K_y(p)} \cdot d(x,y)$$

and mention just another way of expressing the cost function:

$$f^G(p) = \sum_{y \in \mathcal{K}} \frac{K_y(p)}{K} \bar{f}_y(p) \ . \tag{15}$$

We set $\bar{p}_y(x) = p(y|x)/K_y(p)$ and observe that $\sum_{x \in \mathcal{K}} \bar{p}_y(x) = 1$. In the sequel, we set $\gamma = e^{-\varepsilon}$. Let $f(\ell)$ be the function given by (11).

Claim 2. *If $y \in \bar{\mathcal{K}}_1$, then $\bar{f}_y(p) \geq f(\ell)$.*

The proof of Claim 2 is quite simple and hence omitted. In view of (15) and in view of the obvious identity

$$\sum_{y \in \bar{\mathcal{K}}_1} \frac{K_y(p)}{K} = P(\mathcal{K} \times \bar{\mathcal{K}}_1) = 1 - P(\mathcal{K} \times \mathcal{K}_1) \ ,$$

the above claim, in combination with Lemma 6, immediately implies the following second lower bound on the cost function:

$$\begin{aligned} f^G(p) &\geq (1 - P(\mathcal{K} \times \mathcal{K}_1)) \cdot f(\ell) & (16) \\ &> (1 - P(\mathcal{K} \times \mathcal{K}_1)) \cdot \frac{2\gamma}{1-\gamma^2} \cdot \left(1 - \gamma^\ell - \ell\gamma^\ell(1-\gamma)\right) & (17) \\ &\geq (1 - P(\mathcal{K} \times \mathcal{K}_1)) \cdot \frac{2\gamma}{1-\gamma^2} \cdot \left(1 - (s+1)e^{-s}\right) \ , & (18) \end{aligned}$$

where the final inequality is valid provided that $\ell \geq \frac{s}{1-\gamma}$. If $P(\mathcal{K} \times \mathcal{K}_1) \geq 1/s$, we may invoke (14) and conclude that

$$f^G(p) \geq \left(\frac{1}{s} - \frac{4\ell}{K}\right) \cdot \frac{s}{1-\gamma} = \left(1 - \frac{4s\ell}{K}\right) \cdot \frac{1}{1-\gamma} \quad .$$

Otherwise, if $P(\mathcal{K} \times \mathcal{K}_1) < 1/s$, we may invoke (18) and conclude that

$$f^G(p) > \frac{2\gamma}{1-\gamma^2} \cdot \left(1 - \frac{1}{s}\right) \cdot \left(1 - (s+1)e^{-s}\right) \quad .$$

We can summarize this discussion as follows.

Theorem 2. *Let $G = (\mathcal{K}, E)$ be a path of length $K-1$. Suppose that $s \geq 1$, $0 < \gamma < 1$, $\ell \geq \frac{s}{1-\gamma}$ and $K \geq 4\ell$. Then,*

$$\mathrm{OPT}[1] \geq \frac{1}{1-\gamma} \cdot \min\left\{1 - \frac{4s\ell}{K} \ , \ \frac{2\gamma}{1+\gamma} \cdot \left(1 - \frac{1}{s}\right) \cdot \left(1 - (s+1)e^{-s}\right)\right\} \quad .$$

Corollary 3. *With the same notations and assumptions as in Theorem 2, the following holds. If $s \geq 2$ and $K \geq \frac{s^2\ell(1+\gamma)}{\gamma}$, then*

$$\mathrm{OPT}[1] \geq \frac{2\gamma}{1-\gamma^2} \left(\left(1 - \frac{1}{s}\right) \cdot 1 - (s+1)e^{-s}\right) \quad .$$

Proof. For $s \geq 2$ and $K \geq \frac{s^2\ell(1+\gamma)}{\gamma}$ the minimum in Theorem 2 is taken by the second of the two possible terms. □

We would like to show that the parameter vector $(p(y|x))$ which represents the exponential mechanism comes close to optimality. To this end, we need an upper bound on $f^G(p)$. In a first step, we determine an upper bound on the cost induced by the exponential mechanism which makes $p(y|x)$ proportional to $\gamma^{d(x,y)} = \exp(-\varepsilon d(x,y))$. This mechanism might achieve 2ε-differential privacy only. In a second step, we determine an upper bound on the cost induced by the ε-differentially private exponential mechanism which makes $p(y|x)$ proportional to $\gamma^{d(x,y)/2} = \exp(-\varepsilon d(x,y)/2)$. But let's start with the first step.

Lemma 7. *Suppose that the graph $G = (\mathcal{K}, E)$ forms a path of length $K-1$. If p is determined by the 2ε-differentially private exponential mechanism which makes $p(y|x)$ proportional to $\gamma^{d(x,y)}$, then*

$$f^G(p) < \frac{2\gamma}{1-\gamma^2} \quad .$$

Note that this is optimal asymptotically, i.e., when K and the slack parameters ℓ, s in Corollary 3 approach infinity. The proof of Lemma 7 is quite simple and hence omitted. An application of Corollary 3 and of Lemma 7 (with $\gamma^{1/2} = \sqrt{\gamma} = e^{-\varepsilon/2}$ at the place of $\gamma = e^{-\varepsilon}$) immediately leads to the following result:

Corollary 4. *Suppose that the graph $G = (\mathcal{K}, E)$ forms a path of length $K - 1$, $s \geq 2$ and $K \geq \frac{s^2 \ell(1+\gamma)}{\gamma}$. If p is determined by the ε-differentially private exponential mechanism which makes $p(y|x)$ proportional to $\gamma^{d(x,y)/2} = \exp(-\varepsilon d(x,y)/2)$, then*

$$\frac{\mathrm{OPT}^G[1]}{f^G(p)} \geq \frac{\gamma(1-\sqrt{\gamma}^2)}{\sqrt{\gamma}(1-\gamma^2)} \cdot \left(1 - \frac{1}{s}\right) \cdot \left(1 - (s+1)e^{-s}\right)$$
$$\geq \frac{\sqrt{\gamma}}{1+\gamma} \cdot \left(1 - \frac{1}{s}\right) \cdot \left(1 - (s+1)e^{-s}\right) \quad .$$

Note that $\frac{\sqrt{\gamma}}{1+\gamma}$ is close to $1/2$ if γ is close to 1.

6 Worst-case Optimality: Sorting Function

In this section we briefly discuss a scenario where the exponential mechanism is optimal in terms of the worst-case error. More specifically, we consider the problem of publishing the output of the sorting function under differential privacy. Similarly to Example 1 in Section 3, we assume that a database $\mathcal{D}$ is associated with a vector $v \in \mathbb{R}^T$, and that neighboring databases lead to values $v, v' \in \mathbb{R}^T$ such that $\|v - v'\|_1 \leq 2$. For $r \leq T$, the *sorting function* $\pi\colon \mathbb{R}^T \to \mathbb{R}^r$ is defined as follows. For every $v \in \mathbb{R}^T$, take a permutation σ of $1, \ldots, T$ such that $v_{\sigma(1)} \geq \ldots \geq v_{\sigma(T)}$ and define $\pi(v) = (v_{\sigma(1)}, \ldots, v_{\sigma(r)})$. For instance, v may be a frequency list from a password dataset $\mathcal{D}$ and the sorting function applied to v would then return the frequency of the r most chosen passwords in the dataset. In this case, the sorting function π is actually defined over $\mathbb{N}^T$, and inputs can be thought of as histograms. Recent works [3,2] focus on the problem of releasing the whole list of password frequencies under differential privacy, i.e, for $r = T$. Here, we extend the analysis to a more general framework, where r can be arbitrary and the sorting functions are not restricted to histograms.

We first present a lower bound on the minimax risk (under the L_1-norm) that any differentially private mechanism must incur when releasing the output of the sorting function. The omitted proof is based on an application of Assouad's lemma. We underline that Theorem 3 is not entirely original, but carries over and extends a result that appears in a paper currently under review [1].

Theorem 3. *Let $\varepsilon \leq 1/8$. Then, any ε-differentially private mechanism for the sorting function $\pi\colon \mathbb{R}^T \to \mathbb{R}^r$, applied to values with L_1-norm upper-bounded by $N \leq T$, must incur the following minimax risks:*

1. *If $N \leq 1 + 1/(4\varepsilon)$, then $R^\star = \Omega(N)$;*
2. *If $N \geq 1/(2\varepsilon)$, then $R^\star = \Omega(\sqrt{N/\varepsilon})$; or*
3. *If $N \geq r(r+1)/(4\varepsilon) + r$, then $R^\star = \Omega(r/\varepsilon)$.*

We are now ready to prove the optimality of the exponential mechanism when it is used to release the output of the sorting function. For $s \in \mathbb{R}^r$, define

$u(v, s) = -\|\pi(v) - s\|_1$. Note that the exponential mechanism instantiated with this utility function corresponds to the Laplace mechanism which adds Laplace noise with parameter $2/\varepsilon$ to the components of $\pi(v)$. It then is straightforward to show that the error introduced is $O(r/\varepsilon)$. Therefore, for sufficiently large values of N, this upper bound matches the corresponding lower bound in Theorem 3, concluding the analysis.

Acknowledgments. The research was supported by the DFG Research Training Group GRK 1817/1.

References

1. Aldà, F., Simon, H.U.: A lower bound on the release of differentially private integer partitions (2017), submitted to Information Processing Letters
2. Blocki, J.: Differentially private integer partitions and their applications (2016), `tpdp16.cse.buffalo.edu/abstracts/TPDP_2016_4.pdf` (visited 08/08/2016)
3. Blocki, J., Datta, A., Bonneau, J.: Differentially private password frequency lists. In: Proceedings of the 23rd Annual Network and Distributed System Security Symposium (2016)
4. Blum, A., Ligett, K., Roth, A.: A learning theory approach to non-interactive database privacy. J. ACM 60(2), 12 (2013)
5. Brenner, H., Nissim, K.: Impossibility of differentially private universally optimal mechanisms. In: Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science. pp. 71–80 (2010)
6. De, A.: Lower bounds in differential privacy. In: Proceedings of the 9th Theory of Cryptography Conference. pp. 321–338 (2012)
7. Duchi, J.C., Jordan, M.I., Wainwright, M.J.: Local privacy and statistical minimax rates. In: Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science. pp. 429–438 (2013)
8. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Proceedings of the 3rd Theory of Cryptography Conference. pp. 265–284 (2006)
9. Dwork, C., McSherry, F., Talwar, K.: The price of privacy and the limits of lp decoding. In: Proceedings of the 39th annual ACM symposium on Theory of computing. pp. 85–94 (2007)
10. Geng, Q., Kairouz, P., Oh, S., Viswanath, P.: The staircase mechanism in differential privacy. IEEE Journal of Selected Topics in Signal Processing 9(7), 1176–1184 (2015)
11. Ghosh, A., Roughgarden, T., Sundararajan, M.: Universally utility-maximizing privacy mechanisms. In: Proceedings of the 41st Annual ACM Symposium on Theory of Computing. pp. 351–360 (2009)
12. Hall, R., Rinaldo, A., Wasserman, L.: Random differential privacy. Journal of Privacy and Confidentiality 4(2), 43–59 (2012)
13. Hardt, M., Talwar, K.: On the geometry of differential privacy. In: Proceedings of the 42nd ACM Symposium on Theory of Computing. pp. 705–714 (2010)
14. Hsu, J., Roth, A., Roughgarden, T., Ullman, J.: Privately solving linear programs. In: International Colloquium on Automata, Languages, and Programming. pp. 612–624. Springer (2014)

15. Kairouz, P., Oh, S., Viswanath, P.: Extremal mechanisms for local differential privacy. In: Advances in Neural Information Processing Systems. pp. 2879–2887 (2014)
16. Kasiviswanathan, S.P., Lee, H.K., Nissim, K., Raskhodnikova, S., Smith, A.: What can we learn privately. SIAM Journal on Computing 40(3), 793–826 (2011)
17. Koufogiannis, F., Han, S., Pappas, G.J.: Optimality of the Laplace mechanism in differential privacy. arXiv preprint arXiv:1504.00065 (2015)
18. McGregor, A., Mironov, I., Pitassi, T., Reingold, O., Talwar, K., Vadhan, S.: The limits of two-party differential privacy. In: Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science. pp. 81–90 (2010)
19. McSherry, F., Talwar, K.: Mechanism design via differential privacy. In: Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science. pp. 94–103 (2007)