Structure and Sensitivity in Differential Privacy: Comparing K-Norm Mechanisms

Jordan Awan & Aleksandra Slavković

Abstract

Differential privacy (DP), provides a framework for provable privacy protection against arbitrary adversaries, while allowing the release of summary statistics and synthetic data. We address the problem of releasing a noisy real-valued statistic vector T, a function of sensitive data under DP, via the class of K-norm mechanisms with the goal of minimizing the noise added to achieve privacy. First, we introduce the *sensitivity space of* T, which extends the concepts of sensitivity polytope and sensitivity hull to the setting of arbitrary statistics T. We then propose a framework consisting of three methods for comparing the K-norm mechanisms: 1) a multivariate extension of stochastic dominance, 2) the entropy of the mechanism, and 3) the conditional variance given a direction, to identify the optimal K-norm mechanism. In all of these criteria, the optimal K-norm mechanism is generated by the convex hull of the sensitivity space. Using our methodology, we extend the objective perturbation and functional mechanisms and apply these tools to logistic and linear regression, allowing for private releases of statistical results. Via simulations and an application to a housing price dataset, we demonstrate that our proposed methodology offers a substantial improvement in utility for the same level of risk.

Keywords: Statistical Disclosure Control, Entropy, Information Theory, Statistical Depth, Stochastic Dominance, Regression

1 Introduction

Statistical Disclosure Limitation (SDL) or Control (SDC) refers to the broad class of methods developed to address the trade-off between limiting the disclosure risk of sharing and publishing sensitive data, while at the same time maintaining the utility and validity of statistical inference (Duncan and Lambert, 1986, 1989; Fienberg et al., 1998). Until recently, the statistical literature on disclosure limitation has built on a probabilistic notion of disclosure as proposed by Dalenius (1977): "If the release of the statistics [T] makes it possible to determine the value [of confidential statistical data] more accurately than is possible without access to [T], a disclosure has taken place." However, while many statistical procedures appear innocuous, Dwork et al. (2017) demonstrate through a survey of potential disclosure risks (i.e., possible attacks on private data) due to release of aggregate statistics, that it can be nontrivial to determine if a disclosure has taken place or not. In fact, while Dalenius' notion of disclosure risk is intuitive, formalizing this definition requires knowing the prior knowledge of the attacker, their side information, as well as their computational power. In contrast, differential privacy (DP), introduced in (Dwork et al., 2006), has emerged as a formal framework to quantify the "privacy level", which provides provable protection against adversaries with arbitrary priors, unlimited side information, and unbounded computational power. Differential privacy guarantees that whether an individual is in a database or not, the results of a DP procedure should be similar in terms of their probability distribution; this guarantee offers a sense of "plausible deniability" and limits the ability of an adversary to infer about any particular individual in the database. The strength of the privacy guarantee is characterized by a real value $\epsilon > 0$, called the privacy-loss budget, where smaller values of ϵ provide a stronger privacy guarantee.

Differential privacy was proposed by the computer science community, but over the last decade an emphasis has been put on linking DP to fundamental statistical concepts in order to expand and improve its applicability. Wasserman and Zhou (2010) showed that satisfying DP for a small value of ϵ guarantees that certain hypotheses an adversary may attempt to test about particular individuals in the dataset have low power. Wasserman and Zhou (2010) also proposed tools for releasing histograms and density estimates under DP. In Dwork and Lei (2009), it was shown that the amount of noise required to privatize a statistic is related to concepts in the field of robust statistics. Since its inception, the DP framework has been expanded to include mechanisms for private releases of principal components (Chaudhuri et al., 2013), model selection (Lei et al., 2016), hypothesis tests (Vu and Slavkovic, 2009; Wang et al., 2015; Gaboardi et al., 2016; Awan and Slavković, 2018; Canonne et al., 2019), confidence intervals (Karwa and Vadhan, 2017), network analysis (Karwa et al., 2016; Karwa and Slavković, 2016), as well as analyses of functional data (Hall et al., 2013; Mirshani et al., 2019; Awan et al., 2019), to list a few. Besides limiting privacy risk, Dwork et al. (2015a) show that the tools of DP can be used to obtain accurate statistical inferences in adaptive data analysis.

However, DP is often criticized for a substantial drop in statistical utility and thus a lack of applicability, especially in finite sample settings. Bun et al. (2018) show that there is a significant sample complexity cost to answering a set of queries under DP, in terms of the dimension of the database. On the other hand, holding the dimension of the individual's data fixed, Smith (2011) shows that under mild assumptions, asymptotically efficient estimators

are achievable under DP. While it is encouraging that asymptotically, DP mechanisms can perform as well as non-private methods, practical implementations of DP mechanisms often suffer from a considerable drop in performance even for moderate sample sizes. For example, Fienberg et al. (2010) criticized DP mechanisms that release privatized contingency tables for producing unacceptably inaccurate results, and Vu and Slavkovic (2009) show that significant adjustments in the sample size and the sampling distribution of the private test statistic are required to conduct the simplest of hypothesis tests. Therefore, there is a need to not only develop DP tools for more statistical problems, but to optimize existing DP tools to improve their practical accuracy on finite sample data problems. In this paper, we address the problem of releasing a noisy real-valued statistic vector T, a function of sensitive data under DP, via the class of K-norm mechanisms with the goal of minimizing the noise added to achieve privacy, and optimizing the use of the privacy-loss budget ϵ for a fixed statistic and sample size n.

While the computer science community has produced finite sample complexity bounds for many problems (i.e. Hardt and Talwar, 2010; Bun et al., 2018; Cai et al., 2017; Acharya et al., 2018; Canonne et al., 2019 for linear queries, identity testing, and hypothesis testing, respectively, to name a few), there is a much more limited literature on optimizing DP mechanisms for a fixed sample size. Ghosh et al. (2012) show that for any count query, the *geometric mechanism* simultaneously minimizes the expected value of a wide variety of loss functions. Geng and Viswanath (2015) developed a *staircase mechanism* which they show maximizes utility for real-valued statistics, with a focus on ℓ_1 and ℓ_2 error. Geng and Viswanath (2013) show that for two-dimensional real-valued statistics, a correlated multi-dimensional staircase mechanism minimizes the expected ℓ_1 loss. Wang et al. (2014) show that for a database of one person, the minimum-entropy mechanism (with a differentiable density) is Laplace. Wang (2017) shows that for d-dimensional databases with sensitivity based on the ℓ_1 norm, the minimum-entropy mechanism is to add independent Laplace noise to each coordinate. Awan and Slavković (2018) developed uniformly most powerful DP hypothesis tests for Bernoulli data based on the *Tulap distribution*, a variant of the staircase mechanism. Furthermore, Karwa et al. (2016) provide a method of obtaining maximum likelihood estimates for exponential graph models, which correct for the bias introduced by standard DP methods.

The original and most common method of achieving DP for the release of an m-dimensional real statistic vector T is through the Laplace mechanism, which adds independent Laplace random variables to each entry of T before releasing. The Laplace mechanism can be generalized to the K-norm mechanisms introduced by Hardt and Talwar (2010), a family of unbiased mechanisms (mean is the non-private T) which are the focus of this paper. The K-norm mechanisms (here on abbreviated as K-mech) are a family of additive mechanisms determined by the choice of a norm on $\mathbb{R}^m$. Using the ℓ_1 norm leads to the Laplace mechanism which is popularly used (e.g., see Dwork et al. (2006); Smith (2011); Zhang et al. (2012); Yu et al. (2014)). In Chaudhuri and Monteleoni (2009), Chaudhuri et al. (2011), Kifer et al. (2012), Song et al. (2013), and Yu et al. (2014), the ℓ_2 norm variant of the K-mech is used in applications of empirical risk minimization. The ℓ_∞ norm variant of the K-mech is proposed in Steinke and Ullman (2017), as a mechanism to optimize the worst case error when releasing one-way marginals of a high-dimensional binary database. Hall (2012) developed a minimax-optimal K-mech for density estimation, and Xiao and Xiong (2015) study and im-

plement K-mechs related to two-dimensional polytopes for the application of location data. With such a large class of mechanisms, it is natural to ask which K-mech maximizes the utility of the output. Hardt and Talwar (2010) began this comparison by studying K-mechs to release linear statistics, with the goal of optimizing the sample complexity, as the sample size and dimension of T increases. However, the restriction to linear statistics limits the types of problems that can be tackled, and simplifies the geometric problem. In contrast, our goal is to choose the best K-mech to optimize performance for an arbitrary fixed statistic T and sample size, optimizing finite sample statistical utility. By allowing non-linear statistics, we are able to accommodate a larger set of applications; this extension to non-linear statistics also results in more complex geometric structures, which we explore.

Our Contributions The objective of this paper is to optimize the performance of the K-norm mechanisms, for a given arbitrary statistic at a fixed sample size. To do this, we first introduce the new geometric notion, *sensitivity space* S_T of T, related to the *sensitivity* of T (i.e., the amount that T can vary by changing one person's information in the dataset). When releasing a noisy version of a statistic T under ϵ-DP, the amount of noise required is related to the sensitivity of T and an imprecise estimate of the sensitivity, which does not carefully consider the structure of T, can amplify the loss in statistical utility. The proposed sensitivity space allows for the rigorous theoretical and practical comparison of K-Mechs, with the goal of minimizing the magnitude of noise introduced to satisfy ϵ-DP. It also generalizes the polytopes used in Hardt and Talwar (2010), which are called *sensitivity polytope* in Dwork et al. (2014a, 2015b); Nikolov (2015); Kattis and Nikolov (2016), and *sensitivity hull* in Xiao and Xiong (2015); Xiao et al. (2017), all of which are designed for linear statistics. When extended to nonlinear statistics however, the sensitivity space need not be a polytope, and can take a wider variety of forms, which are mathematically more complex. Furthermore, the generalization to arbitrary statistics allows the K-mechs to be implemented within many existing DP mechanisms, and by choosing the K-mech carefully, we can significantly improve the performance of these mechanisms for many statistical problems, such as regression and empirical risk minimization.

In order to identify the optimal K-norm mechanism for a fixed statistic and sample size, we propose a novel framework of comparing the mechanisms that relies on measuring the sensitivity space. The framework consists of proposing three theoretical perspectives of comparing the K-norm mechanisms and linking them with two optimal decision rules. The three perspectives, (1) a multivariate version of stochastic dominance, (2) the entropy of the mechanism, and (3) the conditional variance given a unit direction, each result in one of two stochastic orderings on the K-mechs, which we call the *containment order* and the *volume order*. We show that in all of these criteria, the optimal K-norm mechanism is generated by the convex hull of our proposed sensitivity space, a fundamental result with which we generalize previous results in the literature such as the use of the ℓ_∞ norm in Steinke and Ullman (2017)). The first method of comparison is a novel stochastic ordering which is a multivariate extension of stochastic dominance (Quirk and Saposnik, 1962), motivated by notions of statistical depth (Mosler, 2013), which could be of interest outside of privacy as an alternative stochastic ordering on distributions (Zuo and Serfling, 2000b). Second, we compare the K-mechs in terms of the entropy of the noise-adding distribution. We show that the entropy of a K-mech is determined by the volume of its corresponding norm ball. Finally, we compare K-mechs in terms of their conditional variance given a unit direction, which we

show is based on the containment of the norm balls. This work offers both geometric insight into the mechanisms as well as a simple decision criteria to choose between mechanisms.

Using our new methodology, we also extend the commonly used objective perturbation and functional mechanisms to permit arbitrary K-mechs, allowing for the application of our techniques to optimize the finite sample performance of these mechanisms. Objective perturbation (Chaudhuri and Monteleoni, 2009; Chaudhuri et al., 2011; Kifer et al., 2012) and functional mechanism (Zhang et al., 2012) are highly influential works which have had a large impact on the field of differential privacy, and are among the state of the art mechanisms for regression problems[1]. We illustrate how our theoretical methodology applies to the problems of logistic and linear regression via objective perturbation and functional mechanism, respectively. Through simulations and a real data application, we demonstrate that by carefully choosing the K-mech at crucial steps, we obtain significant gains in the finite-sample accuracy of these mechanisms for the same level of ϵ, improving the applicability of these mechanisms for real data problems. From another perspective, by optimizing the performance we are able to provide the same level of accuracy with a smaller value of ϵ, allowing for better use of the privacy-loss budget to answer other potential statistical queries.

Organization In Section 2, we review the background of DP, introduce the sensitivity space, and demonstrate its connection to the K-norm mechanisms. We explore an example in Section 2.1 to demonstrate the mathematical structure of the sensitivity space for nonlinear statistics. In the main section, Section 3, we propose three methods to compare the K-norm mechanisms. In Subsection 3.1 we propose a stochastic partial ordering of the K-mechs, which in Subsection 3.1.1 we show is connected to concepts in statistical depth. In Subsection 3.2 we derive the entropy of a K-mech in terms of the volume of its corresponding norm ball. In Subsection 3.3 we prove that the conditional variance of the mechanism is optimized based on the containment of the norm balls. In Subsection 3.3, we show that under each of our criteria, the optimal mechanism is generated by the convex hull of the sensitivity space.

In Section 4, we extend the objective mechanism to allow for arbitrary K-mechs. We apply objective perturbation to the problem of logistic regression in Subsection 4.1, and derive the sensitivity space for this problem. In Subsection 4.2, we demonstrate through simulations that choosing the K-mech based on our criteria substantially improves the accuracy of the private logistic regression. In Section 5, we modify the functional mechanism to allow for arbitrary K-mechs, with a focus on linear regression. For this problem the sensitivity space can be written in closed form, allowing us to implement the optimal K-mech exactly. We demonstrate the finite-sample utility gains of our approach in Subsections 5.2 and 5.3 through simulations and a real data example.

We end with concluding remarks and discussion in Section 6. For convenience, we collect algorithms to sample from several K-mechs in the appendix, Subsection 7.1. All proofs and some technical details are postponed to Subsection 7.2.

Throughout the paper, we aim to emphasize which results are our contribution and which are from the previous literature as follows: Any result/definition with a name and no citation

[1]In our preliminary simulations on DP regression mechanisms, we found that objective perturbation and functional mechanism were among the top performing algorithms for logistic and linear regression, respectively, especially for moderate sample sizes.

is original to the present paper, and those with a citation are of course from the literature. Some minor definitions have no name or citation, but are common concepts in the fields of Statistics and Mathematics.

2 Differential Privacy and Sensitivity Space

In this section, we review the necessary background on differential privacy and propose the new concept that we refer to as the *sensitivity space* of a statistic T, which extends the concepts of sensitivity polytope and sensitivity hull to the setting of nonlinear statistics T.

Differential privacy (Dwork et al., 2006), provides a framework for a strong provable privacy protection against arbitrary adversaries while allowing the release of some statistics and potentially synthetic data. It requires the introduction of additional randomness into the analyses such that the distribution of the output does not change substantially if one person were to be in the database or not. A non-technical introduction to DP can be found in Nissim et al. (2017), and a comprehensive introduction can be found in Dwork et al. (2014b).

Before we state the definition of DP, we recall the Hamming distance, which we use to measure the similarity of two databases. The definition of DP can easily be modified to allow for alternative metrics on the space of databases.

Definition 2.1. Let $X, Y \in \mathscr{X}^n$ for any space $\mathscr{X}$. The *Hamming distance* between X and Y is $\delta(X, Y) = \#\{i \mid X_i \neq Y_i\}$, the number of entries where X and Y differ.

Definition 2.2 (Dwork et al., 2006)**.** Let $X = (X_1, \ldots, X_n) \in \mathscr{X}^n$. For $\epsilon > 0$, a mechanism (random function) $M : \mathscr{X}^n \to \mathscr{Y}$ satisfies ϵ-Differential Privacy (ϵ-DP) if for all measurable sets B, and all pairs of databases X and X' such that $\delta(X, X') = 1$, we have that

$$P(M(X') \in B \mid X') \leq \exp(\epsilon) P(M(X) \in B \mid X).$$

Let's take a moment to discuss the cast of characters in Definition 2.2. The value ϵ is called either the privacy parameter, or the privacy-loss budget. Smaller ϵ corresponds to more privacy, but as ϵ approaches infinity there is no privacy guarantee. We think of X_i as the information provided by individual i, so $\mathscr{X}$ is the set of possible observations from one individual. We make no assumptions about the nature of $\mathscr{X}$: while it is common for $\mathscr{X}$ to be a subset of $\mathbb{R}^m$, $\mathscr{X}$ could also be a set of networks, survey responses, or any other data structure. We call $\mathscr{Y}$ the output space, which contains the possible values our statistic of interest can take on. In Definition 2.2 we do not place any restrictions on $\mathscr{Y}$, however in this paper we focus on $\mathscr{Y} = \mathbb{R}^m$ and use Lebesgue measure denote as $\lambda(\cdot)$. Finally the mechanism M is our method of introducing randomness into the output, which is what achieves privacy.

A statistically insightful interpretation of Definition 2.2 was provided by Wasserman and Zhou (2010), connecting it to hypothesis testing. If an adversary wants to test whose data is in the i^{th} entry of X at level α test, Proposition 2.3 assures us that the probability of rejecting the null hypothesis is small.

Proposition 2.3 (Wasserman and Zhou, 2010)**.** *Suppose that $M : \mathscr{X}^n \to \mathscr{Y}$ satisfies ϵ-DP, P is a probability measure on $\mathscr{X}^n$, and $Z = M(X)$ is the released output of the mechanism*

M. Then any level α test which is a function of Z, M, and P, of $H_0 : X_i = u$ versus $H_1 : X_i = v$ has power bounded above by $\alpha \exp(\epsilon)$.

Remark 2.4. Besides Definition 2.2, other formulations of DP have been proposed; see Kifer and Lin (2012) for a formal axiomatization of privacy. For instance, by replacing $\mathscr{X}^n$ with any set, and $\delta(\cdot, \cdot)$ with any metric on that set, one obtains a new notion of DP. Let $\mathscr{X}^* = \{()\} \cup \mathscr{X} \cup \mathscr{X}^2 \cup \cdots$ be the set of all finite tuples with entries in $\mathscr{X}$, where () represents the empty tuple. A popular alternative to Definition 2.2 takes $X \in \mathscr{X}^*$, and $\delta(X, X') = 1$ if X can be obtained from X' by adding or deleting an entry. We will refer to this notion as add/delete-DP, which appears in Dwork et al. (2014b). Note that ϵ-add/delete-DP implies 2ϵ-DP. Other differences are that n is not publicly known, and the mechanism must be well defined on any $X \in \mathscr{X}^*$. While we use Definition 2.2 for concreteness, our theoretical results can be modified to accommodate these other settings.

In this paper, we study mechanisms that add a random vector to a statistic vector T. For such mechanisms, the variance of the random vector, and thus the additonal noise added, must be scaled differently depending on the *sensitivity* of T, which captures the magnitude by which a single individual's data can change the output. The ℓ_1-sensitivity was first introduced in Dwork et al. (2006), but sensitivity can be measured by other norms; e.g., ℓ_2-sensitivity as in Chaudhuri et al. (2011).

As we demonstrate in Sections 4.2, 5.2, and 5.3, the choice of norm for sensitivity calculations can have a significant impact on the performance of DP methods. Next we introduce a key new notion, the *sensitivity space*, whose structure allows us to choose the best norm (See Section 3 for how to evaluate "best") for an arbitrary statistic in $\mathbb{R}^m$. Notions similar to the sensitivity space have played an important role in the theoretical DP literature, but were limited to the study of linear statistics and often considered only ℓ_1 norms on databases. In the case of linear statistics on binary-valued databases, the convex hull of the sensitivity space results in a linear transformation of the L_1 ball, which is referred to as the *sensitivity polytope* (Dwork et al., 2014a, 2015b; Nikolov, 2015; Kattis and Nikolov, 2016). Xiao and Xiong (2015); Xiao et al. (2017) study the setting of discrete location data, where the convex hull of the sensitivity space forms a two-dimensional polytope they call the *sensitivity hull*. In this paper, the proposed sensitivity space is defined for *arbitrary* (not necessary linear) statistics in $\mathbb{R}^m$, allowing for a wider variety of statistical applications, and generalizes the earlier notions of sensitivity polytope/hull. In Subsection 2.1, we explore a simple example which demonstrates the complex geometry of the sensitivity space when applied to non-linear statistics. In Section 3, we show that the optimal K-mech is determined by the convex hull of the sensitivity space. In fact, this is a fundamental result to know[2] in order to improve design of differentially private mechanisms in a formal and principled manner. In Sections 4.1 and 5.1 we demonstrate this by using our more general notion of sensitivity space to optimize DP algorithms for linear and logistic regression.

Definition 2.5 (Sensitivity Space)**.** Let $T : \mathscr{X}^n \to \mathbb{R}^m$ be any function. The *sensitivity space* of T is

$$S_T = \left\{ u \in \mathbb{R}^m \;\middle|\; \begin{array}{c} \exists X, X' \in \mathscr{X}^n \text{ s.t. } \delta(X, X') = 1 \\ \text{and } u = T(X) - T(X') \end{array} \right\}.$$

[2]We thank a reviewer for emphasizing this point.

The sensitivity space consists of all possible differences in the statistic vector T, when computed on two databases differing in one entry. Before introducing the notion of sensitivity, we define *norm balls*, which are the sets in one-to-one correspondence with norms.

Definition 2.6. A set $K \subset \mathbb{R}^m$ is a *norm ball* if K is 1) convex, 2) bounded, 3) absorbing: $\forall u \in \mathbb{R}^m$, $\exists c > 0$ such that $u \in cK$, and 4) symmetric about zero: if $u \in K$, then $-u \in K$.

It is well known that if $K \subset \mathbb{R}^m$ is a norm ball, then we can define a norm $\|\cdot\|_K : \mathbb{R}^m \to \mathbb{R}^{\geq 0}$, given by $\|u\| = \inf\{c \in \mathbb{R}^{\geq 0} \mid u \in cK\}$. We call $\|\cdot\|_K$ the *K-norm*. In fact, any norm $\|\cdot\|$ can be generated this way by taking $K = \{u \mid \|u\| \leq 1\}$.

The sensitivity of a statistic T is the largest amount that T changes when one entry of T is changed. Geometrically, the sensitivity of T is the largest radius of S_T measured by the norm of interest.

Definition 2.7 (K-norm Sensitivity)**.** For a norm ball $K \subset \mathbb{R}^m$, the *K-norm sensitivity* of T is

$$\Delta_K(T) = \sup_{\delta(X,X')=1} \|T(X) - T(X')\|_K = \sup_{u \in S_T} \|u\|_K.$$

For $p \in [1, \infty]$, the *ℓ_p-sensitivity* of T is $\Delta_p(T) = \sup_{u \in S_T} \|u\|_p$. If T is one-dimensional, we simply say the *sensitivity* of T is $\Delta(T) = \sup_{u \in S_T} |u|$.

The family of mechanisms we study in this paper are the K-Norm mechanisms (K-mechs), introduced in Hardt and Talwar (2010). While the focus of Hardt and Talwar (2010) is on linear statistics, and uses a different metric on the input database, they remark that the mechanism is valid in more general settings as well. In Proposition 2.8 and throughout the paper, we note Lebesgue measure as $\lambda(\cdot)$. Recall that the Lebesgue measure of a set S can be interpreted as the volume of S.

Proposition 2.8 (K-Norm Mechanism: Hardt and Talwar, 2010)**.** *Let $X \in \mathscr{X}^n$ and $T : \mathscr{X}^n \to \mathbb{R}^m$. Let $\|\cdot\|_K$ be any norm on $\mathbb{R}^m$ and let $K = \{x \mid \|x\|_K \leq 1\}$ be its unit ball. Call $\Delta_K(T) = \sup_{u \in S_T} \|u\|_K$. Let V be a random variable in $\mathbb{R}^m$, with density $f_V(v) = \frac{\exp(\frac{-\epsilon}{\Delta}\|v\|_K)}{\Gamma(m+1)\lambda(\frac{\Delta}{\epsilon}K)}$, where $\Delta_K(T) \leq \Delta < \infty$. Then releasing $T(X) + V$, satisfies ϵ-DP.*

Since all norms are equivalent in $\mathbb{R}^m$, requiring that $\Delta_K(T) < \infty$ is equivalent to requiring that S_T is bounded.

Since norms are symmetric about zero, any K-mech has mean $T(X)$ and so the K-mechs are a class of unbiased mechanisms, with respect to the randomness introduced for privacy. We refer to the K-mech with norm ℓ_p as the ℓ_p-mechanism (ℓ_p-mech). Note that in the case where the norm is ℓ_1, the mechanism results in adding independent Laplace(Δ/ϵ) noise to each entry of T. So, the K-norm mechanisms can be viewed as a generalization of the Laplace mechanism.

Finally, Proposition 2.9 states that postprocessing cannot increase privacy risk. Postprocessing is both important as a privacy guarantee and as a useful tool to construct complex DP mechanisms (for example, the functional mechanism Zhang et al. (2012)).

Proposition 2.9 (Postprocessing: Dwork et al., 2014b)**.** *Let $X \in \mathscr{X}^n$, $M : \mathscr{X}^n \to \mathscr{Y}$ be a random function, and $f : \mathscr{Y} \to \mathscr{Z}$ be any function. If M is ϵ-DP, then $f \circ M$ is ϵ-DP.*

2.1 Exploring Sensitivity Space

In this subsection, we provide a simple example which illustrates the relation between the sensitivity with respect to different norms, and the sensitivity space. In particular, we demonstrate the complex geometry of the sensitivity space when applied to non-linear statistics.

The K-norm sensitivity of T is often studied as an algebraic object, being the supremum over a set of values. However, we can instead consider how it is geometrically related to the sensitivity space of Definition 2.5. Geometrically, $\Delta_K(T)$ is the radius of the smallest $\|\cdot\|_K$-ball containing S_T. We study the ℓ_p-sensitivity for the following extended example.

Throughout this paper the ℓ_1, ℓ_2, and ℓ_∞-mechs will make frequent appearances. All three of these mechanisms have been seen in the literature and they all have efficient sampling algorithms, which we detail in Section 7.1. On the other hand, while we consider arbitrary K-mechs later in the paper, we acknowledge that in general K-mechs are much harder to implement and sample from. We give a method to sample from arbitrary K-mechs in Subsection 7.1, via rejection sampling.

Example 2.10. Suppose our database is $X = (X_1, \ldots, X_n) \in [-1, 1]^n$, and our statistic of interest is $T(X) = (\sum_{i=1}^n X_i, \sum_{i=1}^n 2X_i^2)$. For this example, the sensitivity space is

$$\begin{aligned} S_T &= \left\{ (u_1, u_2) \in \mathbb{R}^2 \,\middle|\, \begin{array}{rcl} u_1 &=& x_1 - x_2 \\ u_2 &=& 2x_1^2 - 2x_2^2 \end{array} \text{, for some } x_1, x_2 \in [-1, 1] \right\} \\ &= \left\{ (u_1, u_2) \in [-2, 2]^2 \,\middle|\, |u_2| \leq \begin{cases} 2 - 2(1 - u_1)^2 & \text{if } u_1 \geq 0 \\ 2 - 2(u_1 + 1)^2 & \text{if } u_1 < 0 \end{cases} \right\}. \end{aligned}$$

A plot of S_T as a subset of $\mathbb{R}^2$ is shown in Figure 1. For this example, we can work out the ℓ_1, ℓ_2, and ℓ_∞ sensitivities of T exactly:

$$\Delta_1(T) = 3.125, \qquad \Delta_2(T) = 1/4\sqrt{71 + 8\sqrt{2}}, \qquad \Delta_\infty(T) = 2. \tag{1}$$

Of these three, only $\Delta_\infty(T)$ can be computed by inspection. To compute $\Delta_1(T)$ and $\Delta_2(T)$, we solve the calculus problem $\max_{u_1 \in [-2,2]} \left\| \binom{u_1}{u_2} \right\|$, where $u_2 = 2 - 2(u_1 - 1)^2$ (by symmetry of S_T, this is sufficient). The left plot of Figure 1 shows the norm balls $\{u \mid \|u\|_p \leq \Delta_p(T)\}$ for $p = 1, 2, \infty$.

While for this example we are able to compute $\Delta_p(T)$ exactly for $p = 1, 2, \infty$, often in the literature, $\Delta_1(T)$ and $\Delta_2(T)$ are approximated in Equation 2, visualized in the right plot of Figure 1. Such approximations are common; for example, see Zhang et al. (2012) and Yu et al. (2014). In these cases, the ℓ_1 norm is used without considering other options, but by inspecting the right plot of Figure 1 we see that ℓ_∞ is likely a better choice. After developing a formal and systematic method of choosing the best K-mech, we return to this problem in

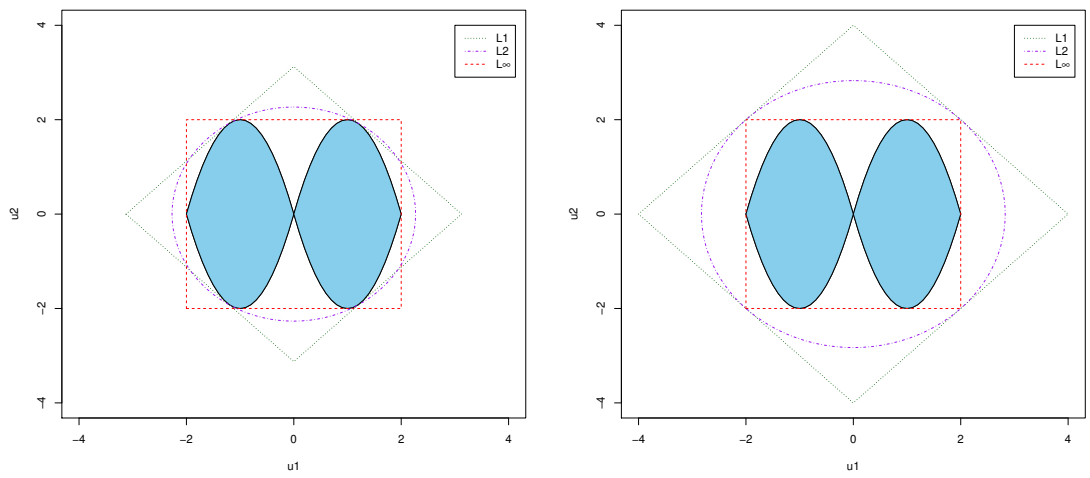

Figure 1: The blue shaded area of both plots is the sensitivity space S_T for Example 2.10. In the left plot, the ℓ_p norm balls of radius $\Delta_p(T)$ as computed in Equation (1), for $p = 1, 2, \infty$ are plotted. In the right plot, the ℓ_p norm balls of radius $\Delta_p(T)$ using the approximations computed in Equation (2), for $p = 1, 2, \infty$ are plotted.

Examples 3.20 and 3.22 to confirm this intuition.

$$\begin{aligned} \Delta_1(T) &= \sup_{\delta(X,X')=1} \|T(X) - T(X')\|_1 \leq \sum_{i=1}^{2} \sup_{\delta(X,X')=1} |T_i(X) - T_i(X')| = 4 \\ \Delta_2(T) &= \sup_{\delta(X,X')=1} \sqrt{\sum_{i=1}^{2} (T_i(X) - T_i(X'))^2} \leq \sqrt{\sum_{i=1}^{2} \sup (T_i(X) - T_i(X'))^2} = \sqrt{8}. \end{aligned} \tag{2}$$

Note that when using these approximations, the norm balls $\{u \mid \|u\|_p \leq \Delta_p(T)\}$ are now larger, as seen in the right plot of Figure 1. On the other hand, in the left plot of Figure 1, we see that when we use the exact sensitivities via Equation (1), none of the norm balls we considered are contained in any of the others. In Section 3, we develop criteria to determine the best norm ball in either scenario. If one norm ball is contained in another, we show in Theorems 3.6 and 3.17 that in a strong sense the smaller norm ball is preferred. If however, neither norm ball is contained in the other, we propose using the norm ball with the smaller volume justified by Theorems 3.15 and Corollary 3.12. In Subsections 4.2, 5.2, and 5.3 we show that choosing the K-mech based on our proposed criteria can have a substantial impact on finite sample utility.

3 Comparing K-Norm Mechanisms

In this section, we develop our main theoretical contributions that enable the comparison of K-norm mechanisms for a given statistic T and sample size n, to determine which mechanism optimizes the finite sample utility. There are several different methods one could use to assess the optimality of a mechanism, for instance in terms of minimizing the expected value of a loss function. However in Section 6, we provide a cautionary example, showing that for such objectives, the optimal mechanism changes for each loss function, and depends on the scaling of the statistic. Instead, we prefer to compare the mechanisms in more intrinsic measures which do not depend on the scaling or coordinate system used. In particular, we propose a novel framework consisting of three methods for comparing the K-mechs in terms of 1) stochastic ordering and statistical depth, 2) entropy, and 3) conditional variance. We show that each of these perspectives results in one of two stochastic orderings on the K-mechs. The first of the two orderings, which we call the *containment order*, compares two K-mechs based on whether one associated norm ball is contained in the other. The other, called the *volume order* is based on the volume of the associated norm balls. The containment order is a partial order, whereas the volume order is a total order which extends the containment order. We show that in both orderings, there is a minimal element which we call the "optimal K-mech," whose corresponding norm is the convex hull of the sensitivity space.

Definition 3.1 (Containment and Volume Orders)**.** Let V and W be two random variables on $\mathbb{R}^m$ with densities $f_V(v) = c\exp(-\frac{\epsilon}{\Delta_K}\|v\|_K)$ and $f_W(w) = c\exp(-\frac{\epsilon}{\Delta_H}\|w\|_H)$. We say that V is preferred over H in the *containment order* if $\Delta_K \cdot K \subset \Delta_H \cdot H$. We say that V is preferred over H in the *volume order* if $\lambda(\Delta_K \cdot K) \leq \lambda(\Delta_H \cdot H)$.

We propose a stochastic ordering which orders random variables based on the containment of certain "level sets," which results in the containment order. We show that this stochastic ordering can also be motivated by concepts in statistical depth; based on the statistical depth literature, we arrive at an alternative stochastic ordering equivalent to the volume order. Another approach to comparing the K-norm mechanisms is based on their entropy, which we show is equivalent to the volume order. Finally, we compare the K-norm mechanisms in terms of their conditional variance, given a unit direction. This comparison also results in the containment order.

In Figure 2, a diagram illustrates our proposed framework, that is the relation between each of our perspectives, and the stochastic orderings associated with them. We can view the items in the top row as theoretical approaches to comparing the K-mechs, and items in the bottom row as decision criteria used to implement the comparison.

3.1 Stochastic Tightness

In this section, we propose a multivariate extension of stochastic dominance, called *stochastic tightness*[3] which orders the K-norm mechanisms based on the geometry of level sets we define and refer to as *concentration sets*, and can be viewed as a multivariate extension of stochastic dominance. The idea is based on showing that the level sets containing probability α of one

[3]Not to be confused with the notion of *tightness* from measure theory.

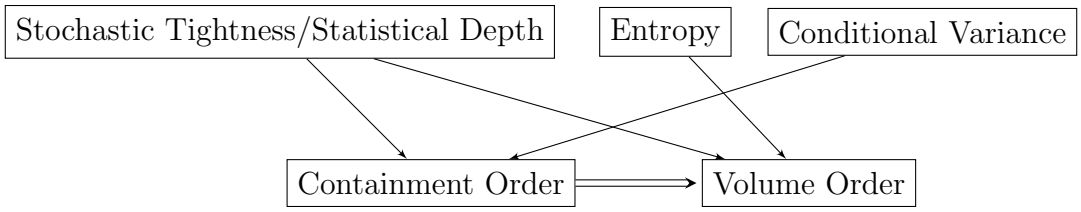

Figure 2: The top row represents the three theoretical perspectives, the bottom represents the two possible decision rules. A standard arrow indicates that a given theoretical perspective justifies the corresponding decision rule. The double arrow indicates that if two mechanisms are ordered with respect to containment, this implies they are ordered with respect to volume.

distribution are always contained in the corresponding level sets of another distribution. In Subsection 3.1.1, we show that these level sets can also be motivated by concepts in the field of statistical depth. When applied to the K-norm mechanisms, we show that this ordering is equivalent to the containment order.

Recall that *stochastic dominance* is a partial order on real-valued random variables, originally proposed in the context of decision theory (Quirk and Saposnik, 1962). A random variable X stochastically dominates Y if $F_X(t) \leq F_Y(t)$ for all $t \in \mathbb{R}$, where $F(\cdot)$ represents the cumulative distribution function. Intuitively if X stochastically dominates Y, then X takes larger values than Y in a strong sense.

There have been several efforts to extend stochastic dominance to 1) compare the dispersion of random variables about a center and 2) allow for the comparison of multivariate distributions. A detailed summary of these various extensions and their connection to statistical depth can be found in Zuo and Serfling (2000b). Here, we introduce *stochastic tightness*, a multivariate stochastic ordering with properties similar to stochastic dominance, which fits into the framework of Zuo and Serfling (2000b). The idea is that the concentration sets, defined in Definition 3.2 serve a similar purpose as the cumulative distribution function. While we use the notion of stochastic tightness in this paper to compare K-norm mechanisms, it is in fact applicable for a wider set of random variables and may be of independent interest in the statistical community.

First, we define the α-concentration set of a random variable.

Definition 3.2 (Concentration Sets)**.** Let X be a unimodal, continuous random variable on $\mathbb{R}^m$ with center a. Assume further that f_X is decreasing along any ray away from the center (i.e. $f(x) \leq f(\alpha x + (1-\alpha)a)$ for all x and all $\alpha \in [0,1]$), and that for all $t > 0$, $P(\{x \mid f(x) = t\}) = 0$. For any $\alpha \in (0,1)$, the set S_X^α is defined to be the smallest (wrt Lebesgue measure) set such that $P(X \in S_X^\alpha) \geq \alpha$. We call S_X^α a *α-concentration set*.

In Definition 3.2, the assumptions imply that the α-concentration set is unique (see property 1) of Lemma 7.3). Note that all of these assumptions are easily verified for any K-mech.

Definition 3.3 (Stochastic Tightness)**.** Let X and Y be random variables on $\mathbb{R}^m$ with center a, which satisfy the assumptions of Definition 3.2. We say that X is *stochastically tighter about a* than Y if for all $\alpha \in (0,1)$, $S_X^\alpha \subseteq S_Y^\alpha$.

Before we investigate the α-concentration sets of the K-norm mechanisms, we require a lemma giving the marginal distribution of the magnitude of a K-norm random variable, sa measured by the K-norm. Conveniently, this works out as a gamma random variable. This result is similar to the decomposition given in Hardt and Talwar (2010), showing that a K-norm random variable can be generated by multiplying a gamma and a uniform random variable. In Kifer et al. (2012), a K-norm random variable is referred to as a multivariate gamma distribution, which this result supports.

Lemma 3.4 (K-norm Marginal Distribution)**.** *For the random variable $V \in \mathbb{R}^m$ with density $f_V(v) \propto \exp(-a\|v\|_K)$, the norm of V is marginally distributed as $\|V\|_K \sim \text{Gamma}(m, a)$.*

In Lemma 3.5, we show that the α-concentration sets for a K-norm random variable are dilations of K. This result is based on two facts: 1) for unimodal and continuous distributions, concentration sets are in 1-1 correspondence with level sets of the density (Casella and Berger, 2002, Theorem 9.3.2), and 2) the density of a K-norm random variable is constant for values with the same K-norm.

Lemma 3.5 (Concentration of K-mech)**.** *For the random variable $V \in \mathbb{R}^m$ with density $f_V(v) \propto \exp(-a\|v\|_K)$, the α-concentration set is*

$$S_V^\alpha = \{v \mid \|v\|_K \leq t\} = tK,$$

where t is the α-quantile of $\text{Gamma}(m, a)$.

Finally, we combine Definition 3.3 and Lemma 3.5 to show that one K-norm mechanism is stochastically tighter than another if its scaled norm ball is contained in the other.

Theorem 3.6 (Stochastic Tightness of K-mechs)**.** *Let K_V and K_W be two norm balls in $\mathbb{R}^m$, let Δ_V and Δ_W be positive real numbers. Define the two random variables V and W on $\mathbb{R}^m$ with densities $f_V(v) \propto \exp\left(\frac{-\epsilon}{\Delta_V}\|v\|_{K_V}\right)$ and $f_W(w) \propto \exp\left(\frac{-\epsilon}{\Delta_W}\|w\|_{K_W}\right)$. The random variable V is stochastically tighter about zero than W if and only if $\Delta_V \cdot K_V \subset \Delta_W \cdot K_W$.*

We see from Theorem 3.6 that the ordering based on stochastic tightness is equivalent to the containment ordering, as defined in Definition 3.1. This gives us our first way of comparing the K-norm mechanisms to determine the optimal mechanism.

Remark 3.7. While in the one-dimensional setting, stochastic dominance implies that the expected value of any increasing objective is maximized, an analogous property does not hold in multivariate settings. In fact, it may be the case that the K-mech is stochastically tighter than the H-mech, and yet many common loss functions are not optimized by K. See a cautionary example in Section 6.

3.1.1 Statistical Depth

Next, we construct a depth function whose depth regions coincide with the α-concentration sets, proposed in the previous section. The α-concentration sets are in fact a special case of *depth regions*, sets determined by a *depth function*. Based on this connection, the statistical

depth literature provides additional justification for the stochastic tightness ordering, and also proposes a second ordering which we show is equivalent to the volume ordering.

First, we review the axioms and terminology of statistical depth, and then construct a depth function whose depth regions agree with the concentration sets of Definition 3.2. The first measure of statistical depth was introduced by Tukey (1975), which is commonly called the half-space depth, or Tukey depth. An alternative to the Tukey depth was introduced by Liu (1988, 1990), and these works provide unifying axioms that characterize statistical depth. Such axioms were further formalized in Zuo and Serfling (2000a). The concepts in statistical depth have also been connected to multivariate notions of order statistics, quantiles, and outlyingnesss measures (Serfling, 2006). See Mosler (2013) for a review of the statistical depth literature and several examples of statistical depth functions.

Definition 3.8 (Depth Function (Zuo and Serfling, 2000a))**.** A *depth function* is a map $D : \mathbb{R}^m \to \mathbb{R}$ such that for any random variable X on $\mathbb{R}^m$ (or a specific sub-class of random variables), the following properties hold:

(A1) (Affine invariance) For any $x \in \mathbb{R}^m$, any full rank $A \in \mathbb{R}^{m\times m}$, and any $b \in \mathbb{R}^m$, we have that $D_{AX+b}(Ax + b) = D_X(x)$.

(A2) (Maximality at the center) If x_0 is the center of X, then $D_X(x_0) = \max_{x\in\mathbb{R}^m} D_X(x)$

(A3) (Linear monotonicity relative to the center) $D_X(x) \leq D_X((1 - \alpha)x_0 + \alpha x)$ for all $\alpha \in [0, 1]$ and all $x \in \mathbb{R}^m$.

(A4) (Vanishing at infinity) $\lim_{\|x\|\to\infty} D_X(x) = 0$.

Intuitively, a *depth function* provides a measure of how close a point in $\mathbb{R}^m$ is to the "center" of a given distribution, with higher values indicating that the point is closer. The first axiom requires that the depth function is invariant under affine transformations, so that the depth measure does not depend on the coordinate system. The third axiom says that the depth function decreases when moving away from the center.

One can use the depth function to draw contours of the distribution illustrating its shape. These contours are called *depth regions*, sets of points with depth greater than a given value. One can also construct depth regions with a specified probability content α as in Definition 3.9, which are conceptually similar to the α-concentration sets of Definition 3.2. Depth regions with probability content α have also been referred to as *depth lifts* (Mosler, 2013).

Definition 3.9 (Depth Region (Zuo and Serfling, 2000c; Mosler, 2013))**.** Given a depth function $D : \mathbb{R}^m \to \mathbb{R}$, the region of depth d for the random variable X is $C_X(d) := \{x \in \mathbb{R}^m \mid D_X(x) \geq d\}$. The *depth region with probability content* α, is $C_X(d(\alpha))$, where $d(\alpha) := \inf\{d \in \mathbb{R} \mid P_X(C_X(d)) \geq \alpha\}$.

Based off of statistical depth, there are two natural stochastic orderings that have been considered in the literature. The first orders distributions based on the containment of depth regions (Mosler, 2013). The second order, proposed by Zuo and Serfling (2000b) orders two distributions based on the volume of the depth regions.

Definition 3.10 (More Dispersed (Mosler, 2013) and More Scattered (Zuo and Serfling, 2000b))**.** Let $D : \mathbb{R}^m \to \mathbb{R}$ be a depth function , and let X and Y be two random variables on $\mathbb{R}^m$. We say that X is *more dispersed* than Y if $C_X(d(\alpha)) \subset C_Y(d(\alpha))$ for all $\alpha \in (0, 1)$. We say that X is *more scattered* than Y if $\lambda(C_X(d(\alpha))) \geq \lambda(C_Y(d(\alpha)))$ for all $\alpha \in (0, 1)$.

In Theorem 3.11, we construct a depth function whose depth regions agree with the α-concentration sets. Verifying the axioms of Definition 3.8 requires a few technical properties of concentration sets, stated in Lemma 7.3, found in the Appendix.

Theorem 3.11 (Depth for Concentration Sets)**.** *Assume that a random variable X on $\mathbb{R}^m$ satisfies all of the conditions in Definition 3.2. Assume further that for all $x \in \mathbb{R}^m$, $\lim_{t\to\infty} f_X(tx) = 0$. Then S_X^α is a depth region with probability content α corresponding to the depth function $D_X(x) = 1 - \inf\{\alpha \mid x \in S_X^\alpha\}$.*

Next we apply the concepts of *more dispersed* and *more scattered* to the K-norm mechanisms with our constructed depth function of Theorem 3.11, and show that these orderings coincide with the containment and volume orders, respectively.

Corollary 3.12 (K-mech More Dispersed/More Scattered)**.** *Let V and W be K and H-norm mechanisms respectively, on $\mathbb{R}^m$. Based on the depth function in Theorem 3.11, W is more dispersed than V if and only if $\Delta_K \cdot K \subset \Delta_H \cdot H$, and W is more scattered than V if and only if $\lambda(\Delta_K \cdot K) \leq \lambda(\Delta_H \cdot H)$.*

Corollary 3.12 provides us with a separate perspective to motivate the containment ordering and offers insight into the notion of stochastic tightness. We also see the volume order for the first time as related to *more scattered*. Thus, both the stochastic tightness and statistical depth perspectives developed here for the K-norm mechanisms can lead to either the containment or volume order decision criteria. In the next subsection, we show that the volume ordering can also be motivated based on the entropy of the K-mechs.

3.2 Entropy

In this section, we compute the entropy of the K-norm mechanisms and show that ordering the K-mechs based on entropy is equivalent to the volume order.

The *entropy* of a random variable is a concept introduced in the field of information theory, and was originally developed to communicate the amount of information that can be sent through a channel, or random variable (Shannon, 1948). For a general introduction to information theory, see (Cover and Thomas, 2012). Roughly speaking, the greater the variability in a random variable, the greater the entropy. By minimizing the entropy of the noise adding distribution, we minimize the amount that the noise can corrupt the non-private signal.

There have been several works connecting the concepts of information theory and differential privacy. Some have studied alternative definitions of DP, phrased in terms of mutual information (Cuff and Yu, 2016; Wang et al., 2016). A few notable works have derived minimum-entropy mechanisms under DP (Wang et al., 2014; Wang, 2017). These works show that for a database of one person, the minimum-entropy mechanism (with a differentiable density) is Laplace. They also show that for d-dimensional databases with sensitivity

based on the ℓ_1 norm, the minimum-entropy mechanism is iid Laplace. Duchi et al. (2013) derive mutual information bounds in the setting of local DP. Rogers et al. (2016) show that an information theoretic concept *max-information* can be used to optimize DP mechanisms for the purpose of hypothesis testing.

Definition 3.13 (Entropy (Shannon, 1948))**.** Let X be a continuous random variable on $\mathbb{R}^m$ with density $f_X(x)$. The *entropy* of X is $H(X) = \mathbb{E}_{f_X}\left(-\log(f_X(X))\right)$. The *conditional entropy* of X given another random variable Y is $H(X \mid Y) = \mathbb{E}_{f_{X,Y}}\left(-\log(f_{X|Y}(X))\right)$, where the expectation is with respect to the joint distribution of X and Y. The *mutual information* of X and Y is $I(X,Y) = \mathbb{E}_{f_{X,Y}} \log\left(\frac{f_{X,Y}(X,Y)}{f_X(X) f_Y(Y)}\right)$.

A useful identity is $I(X,Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$.

Next, we derive a closed-form formula for the entropy of a K-norm mechanism in Proposition 3.14, and show in Theorem 3.15 that ordering K-mechs based on entropy is equivalent to the volume order.

Proposition 3.14 (Entropy of K-mech)**.** *Let V be a random variable on $\mathbb{R}^m$ with density $f_V(v) = (\epsilon/\Delta)^m \frac{\exp(-\frac{\epsilon}{\Delta}\|V\|_K)}{m!\lambda(K)}$. The entropy of V is*

$$H(V) = \log\left(\left(\frac{\Delta \cdot e}{\epsilon}\right)^m m!\lambda(K)\right).$$

From Proposition 3.14, we see that the entropy of a K-norm mechanism is a linear function of $\log\left(\Delta^m \lambda(K)\right)$. So, minimizing the entropy of the mechanism is equivalent to minimizing the volume of $\Delta \cdot K$. Along with Corollary 3.12, we now have a second method of justifying the volume order.

Theorem 3.15 (Entropy Ordering)**.** *Let K_V and K_W be two norm balls in $\mathbb{R}^m$. Consider the random variables V and W on $\mathbb{R}^m$ with densities $f_V(v) \propto \exp(-\frac{\epsilon}{\Delta_V}\|v\|_{K_V})$ and $f_W(w) \propto \exp(-\frac{\epsilon}{\Delta_W}\|w\|_{K_W})$. We have that $H(V) \leq H(W)$ if and only if $\lambda(\Delta_V K_V) \leq \lambda(\Delta_W K_W)$.*

We end this subsection by providing a connection between the entropy of the mechanism and the mutual information between the original statistic and the noisy output. Call T the non-private statistic, V the noise from a K-norm mechanism, and $Z = T + V$ the private output. It is intuitive that we would like to maximize the mutual information between T and Z. However, this quantity depends on the distribution of T, which we assume is unknown. From another perspective, we would like to minimize $I(Z,V)$ which implies that the noise V is not dominating the signal from T. A linear combination of these two objectives can be written in terms of the entropy of T and V:

$$\begin{aligned} I(T,Z) - I(Z,V) &= [H(Z) - H(Z \mid T)] - [H(Z) - H(Z \mid V)] \\ &= H(T) - H(V), \end{aligned}$$

where we use the fact that $H(Z \mid V) = H(T)$. We can view $H(T)$ as an unknown constant. We see that maximizing the objective $I(T,Z) - I(Z,V)$ is equivalent to minimizing $H(V)$.

3.3 Conditional Variance

A natural question is whether there exists a K-mech which minimizes the variance in every direction. In this section, we show that given a direction in $\mathbb{R}^m$, the conditional variance of one K-mech is smaller than another precisely when the associated norm balls are contained.

First, we derive the distribution of a K-mech, conditioned on it lying in a one-dimensional subspace. The distribution is based on the distribution of $\|V\|_K$, developed in Lemma 3.4.

Lemma 3.16 (K-mech Conditional Distribution)**.** *Let $\|\cdot\|_K$ be any norm on $\mathbb{R}^m$. Let V be a random variable with density $f_V(v) \propto \exp(-a\|v\|_K)$. Then*

1. *The random variables $\|V\|_K$ and $\frac{V}{\|V\|_K}$ are independent.*
2. *For any vector $e \in \mathbb{R}^m$ with $\|e\|_2 = 1$, the distribution of $|V^\top e|$ conditional on $V \in \text{span}(e)$, is $\text{Gamma}(m, a\|e\|_K^{-1})$.*

The main result of this subsection, Theorem 3.17 shows that a partial ordering of the K-mechs in terms of their conditional variance is in fact equivalent to the containment order. The proof of Theorem 3.17 uses the conditional distribution developed in Lemma 3.16 and the observation that the variance of $\text{Gamma}(m, \frac{\epsilon}{\Delta_K}\|e\|_K^{-1})$ is minimized by reducing the diameter of $\Delta_K \cdot K$ in the direction of e.

Theorem 3.17 (Conditional Variance of K-mechs)**.** *Let K and H be two norm balls in $\mathbb{R}^m$. Let Δ_K and Δ_H be two positive real numbers. Consider the random variables $V_K, V_H \in \mathbb{R}^m$ drawn from the densities $f(V_K) \propto \exp(\frac{-\epsilon}{\Delta_K}\|V_K\|_K)$ and $g(V_H) \propto \exp(\frac{-\epsilon}{\Delta_H}\|V_H\|_H)$. If $\Delta_K \cdot K \subset \Delta_H \cdot H$, then for all $e \in \mathbb{R}^m$ such that $\|e\|_2 = 1$,*

$$\text{Var}\left(V_K^\top e \mid V_K \in \text{span}(e)\right) \leq \text{Var}\left(V_H^\top e \mid V_H \in \text{span}(e)\right).$$

Theorem 3.17 states that V_K has uniformly smaller variance than V_H, conditional on any direction.

3.4 Optimal K-Norm Mechanism

In Sections 3.1, 3.2, and 3.3, we provided various theoretical perspectives which in turn motivate either the containment or the volume order which provide two decision rules for determining the optimal K-mech. In this section we show that under mild assumptions, the containment and volume order both have the same minimal element, which is determined by the convex hull of the sensitivity space, a fundamental result that allows for a more principled design and evaluation of DP mechanisms.

As noted earlier, under either the containment order or the volume order, we prefer smaller norm balls which contain the sensitivity space S_T. Note that the convex hull of S_T can be expressed as the intersection of all convex sets which contain S_T. Since all norm balls are convex, it follows that the convex hull of S_T is a subset of any norm ball which contains S_T. So, if the convex hull of S_T is a valid norm ball, then it corresponds to the optimal K-norm mechanism under either the containment or volume order. We formalize this observation in Theorem 3.19.

Other works have proposed using the convex hull of the sensitivity space for K-norm mechanisms, but have not formalized it as a fundamental result for the development of DP mechanisms. In Xiao and Xiong (2015), the convex hull is proposed for use in the K-Norm mechanism in the setting of two-dimensional discrete statistics. In Hardt and Talwar (2010), the linear transformations of L_1 balls are the convex hulls of the sensitivity space.

The following lemma establishes when the convex hull leads to a valid norm ball.

Lemma 3.18 (Hull is Norm Ball)**.** *Let $T : \mathscr{X}^n \to \mathbb{R}^m$. Provided that S_T is bounded and $\text{span}(S_T) = \mathbb{R}^m$, then $K_T = \text{Hull}(S_T)$ is a norm ball. So, the norm $\|\cdot\|_{K_T}$ is well defined.*

If S_T is not bounded, then for any norm $\|\cdot\|_K$, the sensitivity $\Delta_K(T)$ is infinite, so no K-norm mechanism can be used to achieve ϵ-DP. If $\text{span}(S_T)$ is a proper subset of $\mathbb{R}^m$, then the entries of $T(X)$ are linearly dependent. So, we can reduce the dimension of T, and recover the removed entries by post-processing.

Theorem 3.19 (Optimal K-mech)**.** *Let $T : \mathscr{X}^n \to \mathbb{R}^m$ such that S_T is bounded and $\text{span}(S_T) = \mathbb{R}^m$. Let $\|\cdot\|_K$ be any norm on $\mathbb{R}^m$, and consider the random variable $V_K \in \mathbb{R}^m$ drawn from the density $f(V_K) \propto \exp\left(\frac{-\epsilon}{\Delta_K}\|\cdot\|_K\right)$. Then V_{K_T} is preferred over V_K in both the containment and volume orders.*

Provided that the conditions of Theorem 3.19 hold, it follows that the convex hull of the sensitivity space gives the K-mech which is least scattered, least dispersed, has minimum entropy, and has minimum conditional variance for all unit directions. Altogether, these properties justify calling this the optimal K-norm mechanism.

3.4.1 Example of Containment and Volume Order

Here, we return to the setting of Example 2.10 and compare various K-mechs with the containment and volume orderings for that problem. We also provide a formula to compute the volume of ℓ_p balls, which appeared in (Wang, 2005), to simplify the calculations needed to apply the volume order.

Example 3.20. Consider the setting of Example 2.10 to determine which mechanism to use based on the containment order. First we note that the convex hull gives a valid norm ball, so this norm is optimal and is written explicitly as K_2 in Section 5.1.

Between ℓ_1, ℓ_2, and ℓ_∞ we have two cases to consider. When using the exact sensitivities in Equation (1), illustrated in the left plot of Figure 1, we see that no norm ball is contained in another. Thus, these K-mechs are incomparable with respect to the containment order, and we cannot determine which mechanism is preferred using this decision criteria. On the other hand, in Example 3.22, we show that the volume order is able to compare these mechanisms. If instead, the sensitivities are approximated as in Equation (2), illustrated in the right plot of Figure 1, we see that the norm balls are strictly contained. Thus, the containment order prefers the mechanisms from best to worst as ℓ_∞, ℓ_2 and finally ℓ_1.

An issue with the containment order is that in higher dimensions, determining containment can be nontrivial. On the other hand, computing volume is relatively simple. In particular, for ℓ_p-balls there is a convenient formula, provided in (Wang, 2005) stated in Proposition 3.21.

Proposition 3.21 (Wang, 2005). *The volume of a unit ℓ_p ball in $\mathbb{R}^m$ is $\frac{2^m\Gamma(1+1/p)^m}{\Gamma(1+m/p)}$.*

Example 3.22. Returning to Example 2.10, using the exact sensitivities of Equation 1 as illustrated in the left plot of Figure 1, the volumes for the ℓ_1, ℓ_2 and ℓ_∞ balls are $\approx$ 19.53, $\approx$ 16.16, and 16, respectively. So, based on the volume order of these three K-norm mechanisms, we prefer the ℓ_∞-mech in this setting. While we only considered ℓ_1, ℓ_2, and ℓ_∞ balls in Example 2.10, by Theorem 3.19 we now know that the optimal K-mech is produced by using the convex hull of the sensitivity space. In fact, we are able to compute the volume of the convex hull of the sensitivity space as $\approx$ 13.33, which offers an even better utility than the ℓ_∞-mechanism.

4 Generalization of Objective Perturbation

In this section we propose a generalization of the objective perturbation mechanism to allow for arbitrary K-norm mechanisms. In Subsection 4.1, we use the techniques of Section 3 to determine the best K-norm mechanisms for use in logistic regression. In Subsection 4.2, we demonstrate through simulations that the choice of mechanism can have a substantial impact on statistical utility.

The objective perturbation mechanism was introduced in Chaudhuri and Monteleoni (2009) for the application of logistic regression. In Chaudhuri et al. (2011), the mechanism was extended to general empirical risk problems, and further extended in Kifer et al. (2012) and Yu et al. (2014).

Algorithm 1 Objective Perturbation as stated in Kifer et al. (2012)

INPUT: $X \in \mathscr{X}^n$, $\epsilon > 0$, a convex set $\Theta \subset \mathbb{R}^m$, a convex function $r : \Theta \to \mathbb{R}$, a convex loss $\hat{\mathscr{L}}(\theta; X) = \frac{1}{n}\sum_{i=1}^n \ell(\theta; x_i)$ defined on Θ such that the Hessian $\nabla^2\ell(\theta; x)$ is continuous in θ and x, $\Delta > 0$ such that $\|\nabla\ell(\theta; x)\|_2 \leq \Delta$ for all $\theta \in \Theta$ and $x \in \mathscr{X}$, and $\lambda > 0$ is an upper bound on the eigenvalues of $\nabla^2\ell(\theta; x)$ for all $\theta \in \Theta$ and $x \in \mathscr{X}$.

1: Set $\gamma = \frac{2\lambda}{\epsilon}$

2: Draw $V \in \mathbb{R}^m$ from the density $f(V; \epsilon, \Delta) \propto \exp(-\frac{\epsilon}{2\Delta}\|V\|_2)$

3: Compute $\theta_{DP} = \arg\min_{\theta\in\Theta} \hat{\mathscr{L}}(\theta; X) + \frac{1}{n}r(\theta) + \frac{\gamma}{2n}\theta^\top\theta + \frac{V^\top\theta}{n}$

OUTPUT: θ_{DP}

In Kifer et al. (2012), it is shown that the output of Algorithm 1 satisfies the add/delete formulation of DP, discussed in Remark 2.4. We need to modify the algorithm to satisfy Definition 2.2. Based on the proof in Kifer et al. (2012) we make several observations. First, to have the output satisfy Definition 2.2, we require that $\sup_{x,x'\in\mathscr{X}} \sup_{\theta\in\Theta}\|\nabla\ell(\theta; x) - \nabla\ell(\theta; x')\| \leq \Delta$. This is related to our notion of sensitivity, but with the inclusion of the parameter θ. Next the use of ℓ_2 norm to measure the sensitivity, and its use in the density of step 2, is arbitrary. Yu et al. (2014) note that ℓ_1 can be used in place of ℓ_2. In fact, any norm can be used along with its K-mech, and so the decision criteria of Section 3 can be applied. Furthermore, we can reduce the size of γ by taking $\gamma = \frac{\lambda}{e^{\epsilon/2}-1} \leq \frac{2\lambda}{\epsilon}$ (see also Yu et al. (2014)). Finally, to control the trade-off between bias and variance, we can introduce a tuning parameter $0 < q < 1$ and replace $f(V; \epsilon, \Delta) \propto \exp(-\frac{\epsilon q}{\Delta}\|V\|_K)$ and $\gamma = \frac{\lambda}{e^{\epsilon(q-1)}-1}$. In Algorithm 1, q is fixed at 1/2. Incorporating these observations, in Algorithm 2, we propose a *generalized objective perturbation mechanism*.

Algorithm 2 Extended Objective Perturbation

INPUT: $X \in \mathscr{X}^n$, $\epsilon > 0$, a convex set $\Theta \subset \mathbb{R}^m$, a convex function $r : \Theta \to \mathbb{R}$, a convex loss $\hat{\mathscr{L}}(\theta; X) = \frac{1}{n}\sum_{i=1}^n \ell(\theta; x_i)$ defined on Θ such that the Hessian $\nabla^2\ell(\theta; x)$ is continuous in θ and x, $\Delta > 0$ such that $\sup_{x,x' \in \mathscr{X}} \sup_{\theta\in\Theta} \|\nabla\ell(\theta; x) - \nabla\ell(\theta; x')\|_K \leq \Delta$ for some norm $\|\cdot\|_K$, $\lambda > 0$ is an upper bound on the eigenvalues of $\nabla^2\ell(\theta; x)$ for all $\theta \in \Theta$ and $x \in \mathscr{X}$, and a real value $0 < q < 1$.

1: Set $\gamma = \frac{\lambda}{\exp(\epsilon(q-1))-1}$

2: Draw $V \in \mathbb{R}^m$ from the density $f(V; \epsilon, \Delta) \propto \exp(-\frac{\epsilon q}{\Delta}\|V\|_K)$

3: Compute $\theta_{DP} = \arg\min_{\theta\in\Theta} \hat{\mathscr{L}}(\theta; X) + \frac{1}{n}r(\theta) + \frac{\gamma}{2n}\theta^\top\theta + \frac{V^\top\theta}{n}$

OUTPUT: θ_{DP}

Theorem 4.1 (Extended Objective Perturbation)**.** *The output of Algorithm 2 satisfies ϵ-DP.*

The proof of Theorem 4.1 mimics the proof in Kifer et al. (2012), and can be found in the Appendix.

4.1 Logistic Regression via Objective Perturbation

In this subsection, we apply the objective perturbation mechanism from Algorithm 2 to the problem of logistic regression. We detail the sensitivity space for this problem, and compare the ℓ_1, ℓ_2 and ℓ_∞ mechanisms based on the containment and volume orderings of Section 3.

Our setup is as follows: we observe $X_{ij} \in [-1, 1]$ and $Y_i \in \{0, 1\}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$ [4]. We take our loss function to be the negative log-likelihood of the logistic regression model:

$$\hat{\mathscr{L}}(\theta; X, Y) = \frac{1}{n}\sum_{i=1}^n \ell(\theta; X_i, Y_i) = \frac{1}{n}\sum_{i=1}^n \log(1 + \exp(\theta^\top X_i)) - Y_i\theta^\top X_i. \tag{3}$$

The gradient and hessian of ℓ are

$$\nabla\ell(\theta; X_i, Y_i) = \left(\frac{\exp(\theta^\top X_i)}{1 + \exp(\theta^\top X_i)} - Y_i\right) X_i, \qquad \nabla^2\ell(\theta; X_i, Y_i) = \left(\frac{\exp(\theta^\top X_i)}{(1 + \exp(\theta^\top X_i))^2}\right) X_i X_i^\top$$

By inspection, we note that the eigenvalues of $\nabla^2\ell(\theta; x, y)$ are bounded above by $\lambda = \frac{m}{4}$. This bound is tight, by taking $X_i = (1, \ldots, 1)^\top$ and $\theta = (0, \ldots, 0)^\top$.

For objective perturbation, the sensitivity space is slightly different than we defined in Definition 2.5. Instead, we also allow for all values of θ:

$$\begin{aligned} S &= \{u \in \mathbb{R}^m | u = \nabla\ell(\theta; X_1, Y_1) - \nabla\ell(\theta; X_2, Y_2), \text{ s.t. } X_1, X_2 \in [-1, 1], Y_1, Y_2 \in \{0, 1\}, \text{ and } \theta \in \mathbb{R}^m\} \\ &= \left\{ \left(\frac{\exp(\theta^\top X_1)}{1 + \exp(\theta^\top X_1)} - Y_1\right) X_1 - \left(\frac{\exp(\theta^\top X_2)}{1 + \exp(\theta^\top X_2)} - Y_2\right) X_2 \right\}. \end{aligned}$$

Note that $\left(\frac{\exp(\theta^\top x)}{1+\exp(\theta^\top x)}\right) \in [0, 1]$ no matter θ or x. We see that the per entry sensitivity of $\nabla\ell$ is bounded above by 2. So, $\Delta_\infty(\nabla\ell) \leq 2$ and $S \subset [-2, 2]^m$. For $m \geq 2$, we have found via simulations that the set $\{c \in \mathbb{R}^m | \exists k \text{ s.t. } c_i \in \{-2, 2\} \text{ for } i \neq k \text{ and } c_k \in \{-1, 1\}\}$. is contained

[4] It may be necessary to rescale and truncate X such as in Zhang et al. (2012) and Lei et al. (2016).

in S. This suggests that while the ℓ_∞-norm may not be optimal, as m increases ℓ_∞ gets closer and closer to optimal. From this, we get the approximate sensitivities $\Delta_\infty(\nabla\ell) = 2$, $\Delta_2(\nabla\ell) = 2\sqrt{m}$, and $\Delta_1(\nabla\ell) = 2m$. In fact these sensitivity calculations place us in a setting similar to the right plot of Figure 1, where the ℓ_∞, ℓ_2, and ℓ_1 balls are contained, in that order. Thus, by either the containment or the volume ordering, we have by Theorems 3.6, 3.15, and 3.17 that ℓ_∞ is the preferred K-mech of these three options. Furthermore, as the convex hull of the sensitivity space is only slightly smaller than the ℓ_∞ ball, we have that the ℓ_∞-mech is nearly optimal in the sense of Theorem 3.19.

4.2 Logistic Regression Simulations

In this section, we implement Algorithm 2 for logistic regression on simulated data. We know from our analysis in the previous section along with the results of Section 3 that ℓ_∞-mech should outperform the ℓ_1 or ℓ_2 mechanisms. We show through simulations that the performance gains by choosing the ℓ_∞-mech are substantial, demonstrating that our choice of K-norm mechanism improves statistical utility.

Our simulation procedure is described in Algorithm 3. For a DP estimate β_{DP}, we measure its performance as the ℓ_2 distance to the true β: $\|\beta_{DP} - \beta\|_2$. We set $n = 10^4$ and consider $\epsilon \in \{1/64, 1/32, \ldots, 1, 2\}$. The DP methods we implement are ℓ_1-mech with $\Delta_1 = 2m$ and $q = 1/2$, ℓ_2-mech with $\Delta_2 = 2\sqrt{m}$ and $q = 1/2$, and ℓ_∞-mech with $\Delta_\infty = 2$ and $q \in \{1/2, .85\}$. First we compare ℓ_1, ℓ_2, and ℓ_2 with $q = 1/2$ as this is the value used in Chaudhuri and Monteleoni (2009); Chaudhuri et al. (2011); Kifer et al. (2012); Yu et al. (2014). We chose $q = .85$ to show that performance can be further improved by tuning q. Unfortunately, we do not tune q under DP so this limits its current usability.

Algorithm 3 Logistic Regression on Simulated Data

INPUT: ϵ and n

1: Set $\beta = (0, -1, \frac{-1}{2}, \frac{-1}{4}, 0, \frac{3}{4}, \frac{3}{2})$ and $m = 7$
2: **for** each of 100 replicates **do**
3: Draw $X_{ij} \overset{\text{iid}}{\sim} U[-1, 1]$ and $U_i \overset{\text{iid}}{\sim} U[0, 1]$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$
4: Set $Y_i = \begin{cases} 1 & \text{if } U_i < e^{X\beta}/(1 + e^{X\beta}) \\ 0 & \text{otherwise} \end{cases}$
5: **end for**
6: **for** each DP estimate and each replicate (X, Y) **do**
7: Compute DP estimate β_{DP} via Algorithm 2
8: Compute ℓ_2 distance to β: $L_{DP} = \|\beta_{DP} - \beta\|_2$
9: **end for**

OUTPUT: $\text{median}_{\text{replicates}}\{L_{DP}\}$ for each method of DP.

In Figure 3, the x-axis indicates the value of ϵ, and the y-axis is the median ℓ_2 distance between the DP estimates and the true β. In this plot, we see that when we fix $q = 1/2$, ℓ_∞ is better than ℓ_2, which beats ℓ_1; for example, ℓ_∞ saves approximately twice the privacy-loss budget ϵ compared to ℓ_2 in this case, which is particularly important for small values of ϵ. Specifically, the ℓ_∞ mechanism achieves a utility value of approximately 1 at $\epsilon = 1/16$, whereas for ℓ_1 to achieve a similar utility, it requires $\epsilon = 1/8$. Recall that the ℓ_2 is the norm used in Chaudhuri and Monteleoni (2009), Chaudhuri et al. (2011) and Kifer et al. (2012), and ℓ_1 is used in Yu et al. (2014). In Yu et al. (2014), they argue that the ℓ_1 should give better performance than ℓ_2, which contradicts our result here. In their analysis, however

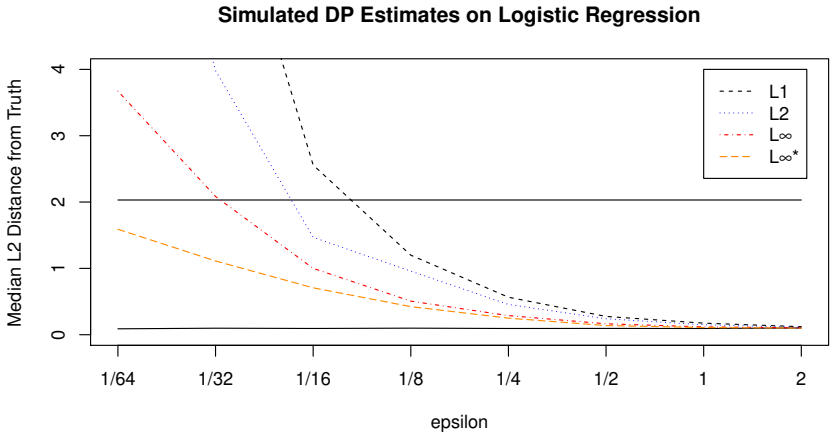

Figure 3: Comparison of ℓ_1, ℓ_2, and ℓ_∞-mechanisms for logistic regression on simulated data, measured by ℓ_2 distance to the true β. The estimates are via Algorithm 2, and the simulation procedure is described in Algorithm 3 with $n = 10^5$. For all estimates, we use $q = 1/2$ except for L^*_∞, which uses $q = .85$. For each ϵ, 100 replicates are used. The upper solid horizontal line indicates the distance between the zero vector and the true β. The lower solid line indicates the distance between the MLE $\hat{\beta}$ and the true β.

they use Δ_1 as an approximation for Δ_2 which hinders the performance of ℓ_2. Instead of either ℓ_1 or ℓ_2, we recommend the ℓ_∞ norm for this application, as the performance gains are substantial. Gains like this could have significant impact on real life applications and usability of DP mechanisms.

We also include the ℓ_∞-mechanism with $q = .85$, labeled as L_{∞^*} in Figure 3. This tuning value q was not chosen under DP, but does demonstrate that utility can be even further improved by considering other values of q. The choice of q under DP is left to future researchers.

5 Linear Regression via Functional Mechanism

In this section, we show that the functional mechanism, a natural mechanism for linear regression, can be easily modified to allow for arbitrary K-norm mechanisms as well. We show that the convex hull of the sensitivity space can be written explicitly in this case, allowing for exact implementation of the optimal K-norm mechanism as determined by Theorem 3.4. In Subsection 5.2, we demonstrate that the optimal K-mech improves the accuracy of the privatized estimates, as measured by the confidence interval coverage. In Subsection 5.3, we show through a real data example that the choice of K-mech reduces the noise introduced in the privatized estimates compared to the non-private estimates.

5.1 Linear Regression Setup

Consider the setting where we have as input X, a $n \times (p+1)$ matrix with left column all 1, and Y a $n \times 1$ vector such that $X_{ij}, Y_i \in [-1, 1]$ for all i, j [5]. We want to estimate β in the model $Y = X\beta + e$, where $e \sim N(0, \sigma^2 I)$. There are many ways of estimating β under DP, such as those discussed in Section 6. Our approach in this section is to sanitize $X^\top X$ and $X^\top Y$ by either ℓ_1-mech, ℓ_∞-mech, or the optimal K-mech, and obtain an estimate of β via post-processing. This approach is similar to the functional mechanism (Zhang et al., 2012), which adds noise to the coefficients of the squared-loss function, before minimizing it. The differences in our approach compared to that in Zhang et al. (2012) are 1) a tighter sensitivity analysis which requires less noise, 2) an extension to any K-mech rather than just ℓ_1-mech, and 3) the use of Equation (4) rather than minimizing the perturbed loss function (which results in more stable estimates). These three differences provide a useful extension of the work in Zhang et al. (2012), resulting in better utility under ϵ-DP.

Let T be the vector of unique, non-constant entries of $X^\top X$ and $X^\top Y$ (T has length $d = [\frac{1}{2}(p+1)(p+2) - 1] + [p+1]$). From the sanitized version of T, we can recover approximations of $X^\top X$ and $X^\top Y$, which we call $(X^\top X)^*$ and $(X^\top Y)^*$ respectively. Then, using the postprocessing property of DP, Proposition 2.9, our DP estimate of β is

$$\hat{\beta}^* = [(X^\top X)^*]^\dagger (X^\top Y)^*, \tag{4}$$

where $A^\dagger$ denotes the Moore-Penrose pseudoinverse of matrix A.

We sanitize T by adding noise from either ℓ_1, ℓ_∞, or the optimal K-mech. We rescale the elements of T so that they all have sensitivity 2. For example, since all $X_{ij} \in [-1, 1]$ the sensitivity of $\sum_i X_{ij}$ is 2, but the sensitivity of $\sum_i X_{ij}^2$ is only 1. So, we replace $\sum_i X_{ij}^2$ with $2\sum_i X_{ij}^2$. The value in rescaling this way is demonstrated in a cautionary example in Section 6. After adding noise, we can divide by 2 to recover our estimate of $\sum_i X_{ij}^2$.

In order to implement the optimal K-mech, we use Algorithm 8 found in Section 7.1 which requires us to sample uniformly from K_T, the convex hull of the sensitivity space

$$S_T = \left\{ u \in \mathbb{R}^m \mid \begin{array}{c} \exists\ \delta\left((X,Y),(X',Y')\right) = 1 \\ \text{s.t. } u = T(X,Y) - T(X',Y') \end{array} \right\}.$$

To understand the geometry of K_T, we consider the two following subproblems:

- The convex hull of the sensitivity space for $(\sum_i X_i, \sum_i 2X_i^2)$, where $X_i \in [-1, 1]$ is

$$K_2 = \left\{ (u_1, u_2) \in [-2, 2]^2 \text{ s.t. } |u_2| \leq \begin{cases} 2 - 2(u_1 - 1)^2, & \text{if } u_1 > 1 \\ 2 - 2(u_1 + 1)^2, & \text{if } u_1 < -1 \end{cases} \right\}.$$

  Note that K_2 is the sensitivity space studied in Example 2.10.

- Suppose we want to release $(\sum_i X_i, \sum_i Y_i, \sum_i X_i Y_i)$ where $X_i, Y_i \in [-1, 1]$. The convex space for this statistic vector is $K_3 = \{(u_1, u_2, u_3) \in [-2, 2]^3 \text{ s.t. } |u_1| + |u_2| + |u_3| \leq 4\}$.

[5] It may be necessary to rescale and truncate X and Y such as in Zhang et al. (2012) and Lei et al. (2016).

For brevity, we omit the arguments that these are indeed the correct convex hulls. Then, K_T is the d-dimensional convex set, which consists of several copies of K_2 and K_3 in different subspaces. This characterization of K_T allows us to determine if a given vector lies in K_T. Using this, we are able to sample from the optimal K-mech via Algorithm 8 in Section 7.

5.2 Linear Regression Simulations

In this section, we measure how close the estimates generated by (4) are to the true β, for each DP mechanism. We consider an estimate close enough to the true β if each entry of the estimate is in the 95% non-private confidence interval (CI) for that entry of β.

The procedure we follow is as described in Algorithm 4. We use point-wise CIs as these are often used by practitioners to determine the significance of coefficients. If the DP estimate is in the CI, one would likely make the same inference using the DP estimate as the MLE.

For our simulations, we set $p = 5$, $n = 10^4$ or $n = 10^6$, and consider $\epsilon \in \{1/16, 1/8, \ldots, 2, 4\}$. The results of these simulations are in Figure 4, where the x-axis denotes varying values of ϵ, and the y-axis measures the proportion of times the estimate $\hat{\beta}^*$ falls in the 95%-CI of β. From these plots, we see that ℓ_∞-mech can reach the performance of ℓ_1-mech with about half the privacy budget. For instance, in the bottom plot of Figure 4 the ℓ_1-mech achieves a fraction of approximately .7 at $\epsilon = 1/2$, whereas the ℓ_∞-mech achieves a similar utility at $\epsilon = 1/4$. This means that choosing ℓ_∞ over ℓ_1-mech results in DP estimates much closer to the true β. On the other hand, the ℓ_∞-mech and the optimal K-mech (derived in Subsection 5.1) perform very similarly. Note that increasing n improves the performance of all methods, but does not change the relative performance of these methods. Additional simulations indicated that the relative performance of the methods is similar for $n = 10^2, 10^3, 10^5$ as well.

Algorithm 4 Simulate Confidence Coverage

INPUT: p, n, ϵ

Set $\beta = (0, -1.5, \ldots, 1.5) \in \mathbb{R}^{p+1}$, where the last p entries are equally spaced.
for each of 200 replicates **do**
    Draw $X^0_{ij} \overset{\text{iid}}{\sim} U[-1, 1]$ for $i = 1, \ldots, n$ and $j = 1, \ldots, p$
    Set $X = [1_n, X^0]$ and Draw $Y \sim N(X\beta, I_n)$
end for
for each replicate (X, Y) **do**
    Compute 95% CI for each of the last p entries of β, based on $\hat{\beta}_{MLE}$
end for
for each DP method and each replicate (X, Y) **do**
    Compute the DP estimate β_{DP} via (4)
    Compute average coverage :
    $C_{DP} = \frac{1}{p} \sum_{i=1}^{p} \#$(entries of β_{DP} in its CI).
end for

OUTPUT: $\frac{1}{200} \sum_{\text{replicates}} C_{DP}$ for each method of DP.

5.3 Linear Regression on Housing Data

In this section we analyze a dataset containing information on 348,189 houses in the San Francisco Bay area, collected between 2003 and 2006. Our response is rent, and the predictors are lot square-footage, base square-footage, location in latitude and longitude, time of

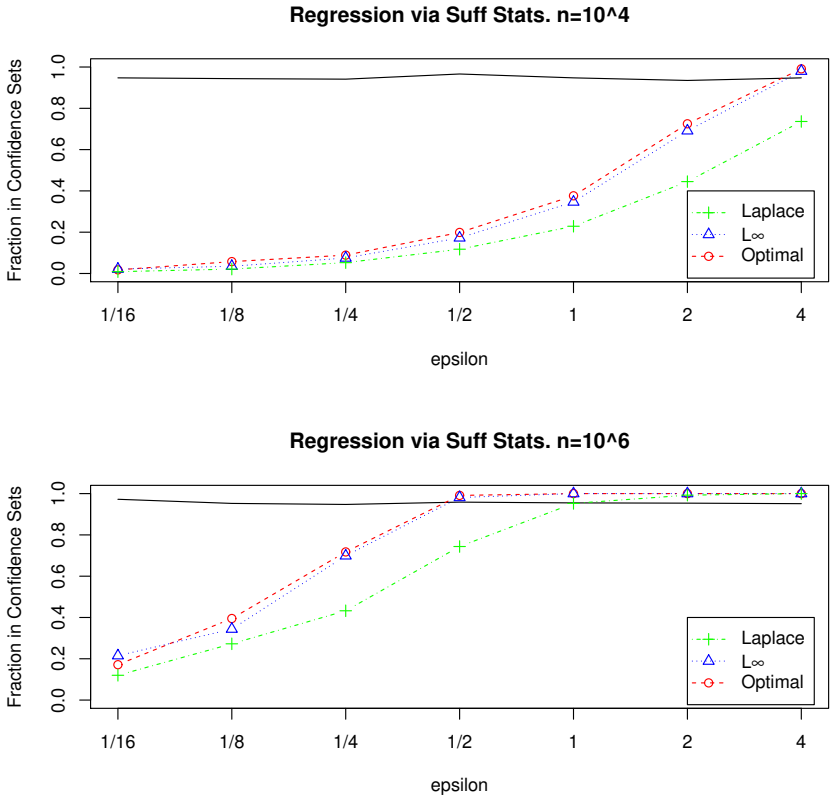

Figure 4: Comparison of ℓ_1-mech, ℓ_∞-mech, and optimal K-mech for linear regression, via Algorithm 4. The estimates used in the above plot are via (4). For all simulations, $p = 5$ and 200 replicates are used. In the top plot, $n = 10^4$ and in the bottom plot, $n = 10^6$. The solid line is how often the true β falls in the confidence intervals, which is $\approx .95$.

transaction, age of house, number of bedrooms, and five indicators for the counties: Alameda, Contra Costa, Marin & San Francisco & San Mateo, Napa & Sonoma, Santa Clara.

To clean the data, we follow a similar procedure as in Lei (2011) and Lei et al. (2016). We remove houses with prices outside of the range 105 to 905 thousand dollars, as well as houses with square-footage larger than 3000. In total, we have one response, 12 predictors, and 235,760 observations. As additional pre-processing, we apply a log-transformation to rent and both measures of square-footage. We then truncate all variables between the 0.0001 and .9999 quantiles and then linearly transform the truncated variables to lie in $[-1, 1]$. This procedure results in well-distributed values in each attribute. We also found that after this pre-processing, the assumptions of the linear model were reasonable.

As described in Subsection 5.1, we form the vector T based on this data, add to it noise from either ℓ_1-mech, ℓ_∞-mech, or the optimal K-mech, and post-process T to get an estimate of the coefficient vector β via (4).

We measure the performance of each DP estimate β_{DP} by its ℓ_2 distance to the MLE estimate β_{MLE}: $\|\beta_{DP} - \beta_{MLE}\|_2$. We give plots of the performance of ℓ_1-mech, ℓ_∞-mech, and

the optimal K-mech under this measure in Figure 5. Each curve is an aggregate over 1000 replications of the DP algorithm. We see in Figure 5 that choosing ℓ_∞-mech over ℓ_1-mech can affect performance about as much as doubling the privacy budget ϵ. For a fixed ϵ, the estimates from ℓ_∞-mech are considerably closer to the MLE and give substantially better estimates than the estimates from ℓ_1-mech. On the other hand, the optimal K-mech does not offer sizeable benefits over ℓ_∞-mech.

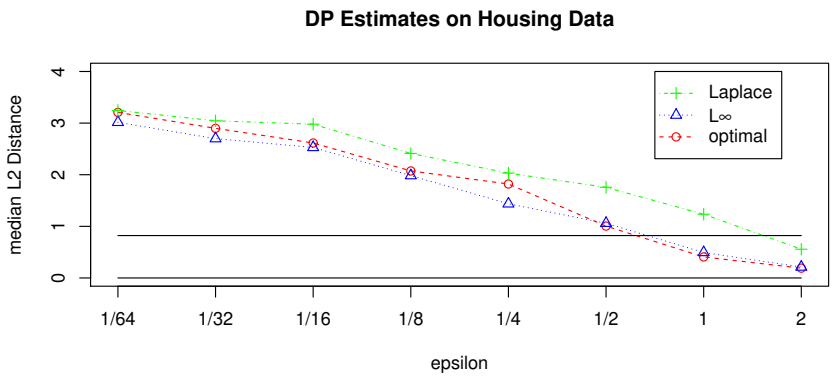

Figure 5: Comparison of ℓ_1-mech, ℓ_∞-mech, and optimal K-mech, calculated via (4) for linear regression on Housing Data, measured by ℓ_2 distance to $\hat{\beta}$. The solid line at height $\approx .82$ is the ℓ_2 distance between the zero vector and $\hat{\beta}$. The lower horizontal line is at height 0. For each ϵ, 1000 replicates are aggregated for each DP mechanism.

Remark 5.1. The ℓ_1 and ℓ_2 sensitivities used in this example are not exact, and it may be possible to improve the performance of the ℓ_1 and ℓ_2 mechanisms slightly by optimizing these sensitivity calculations. Furthermore, the sensitivity space developed in Subsection 5.1 assumed that all coordinates in X can take values between -1 and 1. However, in this example we have indicator variables which are inherently dependent (i.e. if the Alameda indicator variable is active, then the Contra Costa indicator must be inactive). This additional structure implies that the sensitivity space is actually smaller than the generic one developed in Subsection 5.1. So, while all of our mechanisms in this section are valid, it is possible that they could be even further improved by taking these observations into account.

6 Discussion

In this paper, we address the problem of releasing a noisy real-valued statistic vector T, a function of sensitive data under DP, via the class of K-norm mechanisms with the goal of minimizing the noise added to achieve privacy, and optimizing the use of the privacy-loss budget ϵ. We propose a new notion we refer to as the *sensitivity space* to understand the geometric relation between a statistic and its sensitivity. We used the sensitivity space to study the class of K-mechs, a natural extension of the Laplace mechanism. Rather than naively using iid Laplace (the ℓ_1-mech) or any other K-mech, we recommend choosing the

K-mech based on properties of the sensitivity space. To this end we propose three methods of evaluating the K-norm mechanisms in order to identify the optimal one for a fixed arbitrary (linear or non-linear) statistic T and sample size n. We then showed a result fundamental for designing improved differentially private mechanisms: that the convex hull of the sensitivity space results in the optimal K-mech, which is stochastically tightest, has minimum entropy, and minimizes the conditional variance. On the other hand, if two (or more) K-mechs in particular are to be compared, this can be done by either checking for the containment of their associated norm balls, or by comparing the volume of their norm balls. In this case, even if using the convex hull is computationally intractible, we offer a framework as illustrated in Figure 2, which results in simple decision criteria to choose between several candidate K-mechs.

The proposed volume and containment orderings could be of broader statistical interest outside of privacy. In fact, we show that these orderings are connected to *more scattered* and *more dispersed*, which extend the notion of stochastic dominance to multivariate settings Zuo and Serfling (2000b). Furthermore, as we showed in Section 3.1.1 stochastic tightness is closely related to stochastic depths, and may be of interest in that field as well.

Our extensions of objective perturbation and functional mechanism are also significant in the broader differential privacy community. Our modifications emphasize the flexibility that these mechanisms have in tailoring the noise-adding distribution to the problem at hand, and we show that by using the comparison criteria to determine the optimal K-mech, we are able to improve the performance of the output of these mechanisms, in terms of statistical utility. To facilitate the ease of implementing our proposed methods, we provide a method of sampling arbitrary K-mechs in Subsection 7.1, via rejecting sampling, and collect computationally efficient algorithms to sample the ℓ_1, ℓ_2, and ℓ_∞ mechanisms from the literature as well.

While the focus of this paper has been on reducing the noise introduced at a fixed level of ϵ, alternatively our techniques allow one to achieve a desired amount of accuracy with a reduced privacy level ϵ. This saves more of the privacy-loss budget for the computation of additional statistics or other statistical tasks, improving the usability of differentially private techniques. In fact, through the applications of linear and logistic regression, via the functional and objective perturbation mechanisms, we showed that choosing the K-mech based on our proposed criteria allows the same accuracy to be achieved with about half of the privacy-loss budget. The question of determining and setting the privacy-loss budget has come to prominence more recently with advances in DP methodologies and tools, and implementation of formal privacy in large organizations, especially those who must share data more broadly for an array of potential statistical analyses and maintain confidentiality, such as the U.S. Census. Theoretical results and practically implementable solutions such as those presented in this paper help address this fundamental question,

In Section 3, we developed several methods of comparing K-norm mechanisms, and in Sections 4.2, 5.2, and 5.3 we demonstrated that the proposed criteria provide a substantial improvement in practical utility. However, this may not always be the case. In particular, we provide an example below illustrating that even when the norm ball H is contained in K, this does not imply that the H-mechanism outperforms the K-mechanism in terms of the marginal variance or the expected value of the ℓ_1, ℓ_2, or ℓ_∞ loss.

A cautionary Example Consider the example where K, the convex hull of the sen-

sitivity space, is the blue solid rectangle in Figure 6, and H is the red textured diamond which contains K. We assume that the sensitivity of both is one. Then Theorem 3.17 applies in this setting, and we know that the K-norm mechanism has smaller conditional variance than the conditional variance of the H-norm mechanism in every direction. However, the marginal variance of H in the x and y coordinates are approximately 24139.87 and 242.09 respectively, whereas the marginal variance of K in the same directions are 40068.37 and 3.99 using 100,000 samples from both mechanisms. In Table 6, we see that the expected L_∞, L_1 and L_2 loss are all minimized by H rather than K. This example demonstrates a "Simpson's paradox" (Blyth, 1972), where the behavior of conditional variables is very different from the marginal variables

One reason this phenomena occurs is that the scaling in the x and y directions are of very different magnitudes. In terms of the marginal variance, the K-mech has a relatively large reduction in the y direction, but a relatively moderate increase in the x direction. However, in absolute terms, the increase in the x direction dwarfs the reduction in the y-direction.

This same phenomenon could be constructed by instead setting the scales of x and y more equally, but using a loss function that disproportionately penalizes variability in the x direction. Thus, it is important to note that in multivariate settings, without knowing the scaling of the variables and the particular loss function of interest, we cannot guarantee that any K-mech will outperform another in terms of these particular metrics.

On the other hand by scaling each entry of the statistic T equally, as we did in Sections 4.1 and 5.1 we are able to mitigate this problem as we saw in our numerical examples.

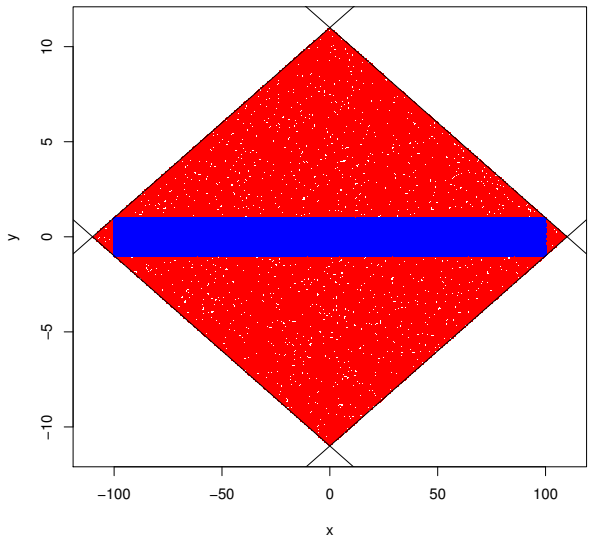

Figure 6: The solid blue rectangle represents the space K and the textured red diamond represents the space H. Note that K is entirely contained in H.

Table 1: Expected loss of the K and H-norm mechanisms, described in Section 2. Monte carlo standard errors are in parentheses.

	ℓ_∞	ℓ_2	ℓ_1
K	150.17(0.418)	150.19(0.418)	151.66(0.420)
H	110.89(0.345)	112.27(0.343)	120.89(0.349)

Future work As the Laplace mechanism is a part of many other mechanisms (i.e., Stochastic Gradient Descent (Song et al., 2013) and Subsample-Aggregate (Smith, 2011)), our methodology can be used to improve finite sample performance of other mechanisms as well. Recently, Reimherr and Awan (2019) developed a new mechanism, which can be viewed as a hybrid of the exponential mechanism and objective perturbation, which they call the K-norm gradient mechanism (KNG). Using the sensitivity space and tools of this paper, we may be able to optimize the performance of KNG, by choosing the optimal norm. Optimizing the performance of these mechanisms allows for improved statistical inference, increased usability of DP methods, and better use of the privacy-loss budget.

References

Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. Differentially private testing of identity and closeness of discrete distributions. In *Advances in Neural Information Processing Systems*, pages 6878–6891, 2018.

Jordan Awan and Aleksandra Slavković. Differentially private uniformly most powerful tests for binomial data. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 4208–4218. Curran Associates, Inc., 2018.

Jordan Awan, Ana Kenney, Matthew Reimherr, and Aleksandra Slavković. Benefits and pitfalls of the exponential mechanismwith applications to hilbert spaces and functional pca. In *Proceedings of the 36th International Conference on International Conference on Machine Learning*, ICML'19, pages 374–384. JMLR.org, 2019.

Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. ISBN 0387310738.

Colin R Blyth. On simpson's paradox and the sure-thing principle. *Journal of the American Statistical Association*, 67(338):364–366, 1972.

Mark Bun, Jonathan Ullman, and Salil Vadhan. Fingerprinting codes and the price of approximate differential privacy. *SIAM Journal on Computing*, 47(5):1888–1938, 2018.

Bryan Cai, Constantinos Daskalakis, and Gautam Kamath. Priv'it: priv ate and sample efficient i dentity t esting. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 635–644. JMLR. org, 2017.

Clément L Canonne, Gautam Kamath, Audra McMillan, Adam Smith, and Jonathan Ullman. The structure of optimal private tests for simple hypotheses. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 310–321. ACM, 2019.

George Casella and Roger L Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

Kamalika Chaudhuri and Claire Monteleoni. Privacy-preserving logistic regression. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 289–296. Curran Associates, Inc., 2009.

Kamalika Chaudhuri, Claire Monteleoni, and D. Sarwate. Differentially private empirical risk minimization. In *Journal of Machine Learning Research*, volume 12, pages 1069–1109, 2011.

Kamalika Chaudhuri, Anand D. Sarwate, and Kaushik Sinha. A near-optimal algorithm for differentially-private principal components. *Journal of Machine Learning Research*, 14(1): 2905–2943, January 2013. ISSN 1532-4435.

Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

Paul Cuff and Lanqing Yu. Differential privacy as a mutual information constraint. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 43–54. ACM, 2016.

T. Dalenius. Towards a methodology for statistical disclosure control. *Statistik Tidskrift*, 15: 429–444, 1977.

John C Duchi, Michael I Jordan, and Martin J Wainwright. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 429–438. IEEE, 2013.

George Duncan and Diane Lambert. Disclosure-limited data dissemination. *Journal of The American Statistical Association*, 81:10–18, 03 1986.

George Duncan and Diane Lambert. The risk of disclosure for microdata. *Journal of Business & Economic Statistics*, 7(2):207–217, 1989. doi: 10.1080/07350015.1989.10509729.

Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing*, STOC '09, pages 371–380, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-506-2. doi: 10.1145/1536414.1536466.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. *Calibrating Noise to Sensitivity in Private Data Analysis*, pages 265–284. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. ISBN 978-3-540-32732-5. doi: 10.1007/11681878_14.

Cynthia Dwork, Aleksandar Nikolov, and Kunal Talwar. Using convex relaxations for efficiently and privately releasing marginals. In *Proceedings of the thirtieth annual symposium on Computational geometry*, page 261. ACM, 2014a.

Cynthia Dwork, Aaron Roth, et al. *The algorithmic foundations of differential privacy*, volume 9. Now Publishers, Inc., 2014b.

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, New York, NY, USA, 2015a. ACM. ISBN 978-1-4503-3536-2. doi: 10.1145/2746539.2746580.

Cynthia Dwork, Aleksandar Nikolov, and Kunal Talwar. Efficient algorithms for privately releasing marginals via convex relaxations. *Discrete & Computational Geometry*, 53(3): 650–673, 2015b.

Cynthia Dwork, Adam Smith, Thomas Steinke, and Jonathan Ullman. Exposed! a survey of attacks on private data. *Annual Review of Statistics and Its Application*, 4(1):61–84, 2017. doi: 10.1146/annurev-statistics-060116-054123.

Stephen E. Fienberg, Udi E. Makov, and Russell J. Steele. Disclosure limitation using perturbation and related methods for categorical data. *Journal of Official Statistics*, Vol. 14(No. 4):485–502, 1998.

Stephen E. Fienberg, Alessandro Rinaldo, and Xiaolin Yang. Differential privacy and the risk-utility tradeoff for multi-dimensional contingency tables. In Josep Domingo-Ferrer and Emmanouil Magkos, editors, *Privacy in Statistical Databases*, pages 187–199, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg. ISBN 978-3-642-15838-4.

Marco Gaboardi, Hyun Lim, Ryan Rogers, and Salil Vadhan. Differentially private chi-squared hypothesis testing: Goodness of fit and independence testing. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2111–2120, New York, New York, USA, 20–22 Jun 2016. PMLR.

Quan Geng and Pramod Viswanath. The optimal mechanism in differential privacy: Multi-dimensional setting. *arXiv preprint arXiv:1312.0655*, 2013.

Quan Geng and Pramod Viswanath. The optimal noise-adding mechanism in differential privacy. *IEEE Transactions on Information Theory*, 62(2):925–951, 2015.

Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. *SIAM Journal on Computing*, 41(6):1673–1693, 2012.

Rob Hall, Alessandro Rinaldo, and Larry Wasserman. Differential privacy for functions and functional data. *Journal of Machine Learning Research*, 14(1):703–727, February 2013. ISSN 1532-4435.

Robert Hall. *New Statistical Applications for Differential Privacy*. PhD thesis, Carnegie Mellon, December 2012.

Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *Proceedings of the Forty-second ACM Symposium on Theory of Computing*, STOC '10, pages 705–714, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0050-6. doi: 10.1145/1806689.1806786.

Vishesh Karwa and Aleksandra Slavković. Inference using noisy degrees: Differentially private β-model and synthetic graphs. *The Annals of Statistics*, 44(1):87–112, 02 2016. doi: 10.1214/15-AOS1358.

Vishesh Karwa and Salil P. Vadhan. Finite sample differentially private confidence intervals. *CoRR*, abs/1711.03908, 2017.

Vishesh Karwa, Pavel N. Krivitsky, and Aleksandra B. Slavković. Sharing social network data: differentially private estimation of exponential family randomgraph models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(3):481–500, 2016. doi: 10.1111/rssc.12185.

Assimakis Kattis and Aleksandar Nikolov. Lower bounds for differential privacy from gaussian width. *arXiv preprint arXiv:1612.02914*, 2016.

D Kifer, A Smith, and A Thakurta. Private convex empirical risk minimization and high-dimensional regression. *Journal of Machine Learning Research*, 1:1–41, 01 2012.

Daniel Kifer and Bing-Rong Lin. An axiomatic view of statistical privacy and utility. *Journal of Privacy and Confidentiality*, 4(1), 2012.

Jing Lei. Differentially private m-estimators. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 361–369. Curran Associates, Inc., 2011.

Jing Lei, AnneSophie Charest, Aleksandra Slavkovic, Adam Smith, and Stephen Fienberg. Differentially private model selection with penalized and constrained likelihood. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 0(0), 2016. doi: 10.1111/rssa.12324.

Regina Y Liu. On a notion of simplicial depth. *Proceedings of the National Academy of Sciences*, 85(6):1732–1734, 1988.

Regina Y Liu. On a notion of data depth based on random simplices. *The Annals of Statistics*, 18(1):405–414, 1990.

Ardalan Mirshani, Matthew Reimherr, and Aleksandra Slavković. Formal privacy for functional data with gaussian perturbations. In *International Conference on Machine Learning*, pages 4595–4604, 2019.

Karl Mosler. Depth statistics. In *Robustness and complex data structures*, pages 17–34. Springer, 2013.

Aleksandar Nikolov. An improved private mechanism for small databases. In *International Colloquium on Automata, Languages, and Programming*, pages 1010–1021. Springer, 2015.

Kobbi Nissim, Thomas Steinke, Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, David R OBrien, and Salil Vadhan. Differential privacy: A primer for a non-technical audience. In *Privacy Law Scholars Conf*, 2017.

James P Quirk and Rubin Saposnik. Admissibility and measurable utility functions. *The Review of Economic Studies*, 29(2):140–146, 1962.

Matthew Reimherr and Jordan Awan. Kng: The k-norm gradient mechanism. *arXiv preprint arXiv:1905.09436*, 2019.

Ryan Rogers, Aaron Roth, Adam Smith, and Om Thakkar. Max-information, differential privacy, and post-selection hypothesis testing. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 487–494. IEEE, 2016.

Robert Serfling. Depth functions in nonparametric multivariate inference. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 72:1, 2006.

Claude Elwood Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

Adam Smith. Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the Forty-third Annual ACM Symposium on Theory of Computing*, STOC '11, pages 813–822, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0691-1. doi: 10.1145/1993636.1993743.

Shuang Song, Kamalika Chaudhuri, and Anand D. Sarwate. Stochastic gradient descent with differentially private updates. In *in Proceedings of the Global Conference on Signal and Information Processing. IEEE*, pages 245–248, 2013.

Thomas Steinke and Jonathan Ullman. Between pure and approximate differential privacy. *Journal of Privacy and Confidentiality*, 7, 2017.

John W Tukey. Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians, Vancouver, 1975*, volume 2, pages 523–531, 1975.

Duy Vu and Aleksandra Slavkovic. Differential privacy for clinical trial data: Preliminary evaluations. In *Proceedings of the 2009 IEEE International Conference on Data Mining Workshops*, ICDMW '09, pages 138–143, Washington, DC, USA, 2009. IEEE Computer Society. ISBN 978-0-7695-3902-7. doi: 10.1109/ICDMW.2009.52.

Weina Wang, Lei Ying, and Junshan Zhang. On the relation between identifiability, differential privacy, and mutual-information privacy. *IEEE Transactions on Information Theory*, 62(9):5018–5029, 2016.

Xianfu Wang. Volumes of generalized unit balls. *Mathematics Magazine*, 78(5):390–395, 2005.

Y. Wang, J. Lee, and D. Kifer. Revisiting Differentially Private Hypothesis Tests for Categorical Data. *ArXiv e-prints*, November 2015.

Yu Wang, Zhenqi Huang, Sayan Mitra, and Geir E Dullerud. Entropy-minimizing mechanism for differential privacy of discrete-time linear feedback systems. In *53rd IEEE Conference on Decision and Control*, pages 2130–2135. IEEE, 2014.

Yu-Xiang Wang. Per-instance differential privacy and the adaptivity of posterior sampling in linear and ridge regression. *stat*, 1050:18, 2017.

Larry Wasserman and Shuheng Zhou. A statistical framework for differential privacy. *JASA*, 105:489:375–389, 2010.

Yonghui Xiao and Li Xiong. Protecting locations with differential privacy under temporal correlations. In *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security*, CCS '15, pages 1298–1309, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3832-5. doi: 10.1145/2810103.2813640.

Yonghui Xiao, Li Xiong, Si Zhang, and Yang Cao. Loclok: Location cloaking with differential privacy via hidden markov model. *Proceedings of the VLDB Endowment*, 10(12):1901–1904, 2017.

Fei Yu, Michal Rybar, Caroline Uhler, and Stephen E. Fienberg. Differentially-private logistic regression for detecting multiple-snp association in gwas databases. In *Privacy in Statistical Databases: UNESCO Chair in Data Privacy, International Conference, PSD 2014, Ibiza, Spain, September 17-19, 2014. Proceedings*, pages 170–184, Cham, 2014. Springer International Publishing. ISBN 978-3-319-11257-2. doi: 10.1007/978-3-319-11257-2_14.

Jun Zhang, Zhenjie Zhang, Xiaokui Xiao, Yin Yang, and Marianne Winslett. Functional mechanism: Regression analysis under differential privacy. *Proc. VLDB Endow.*, 5(11): 1364–1375, July 2012. ISSN 2150-8097. doi: 10.14778/2350229.2350253.

Yijun Zuo and Robert Serfling. General notions of statistical depth function. *Annals of statistics*, pages 461–482, 2000a.

Yijun Zuo and Robert Serfling. Nonparametric notions of multivariate scatter measure and more scattered based on statistical depth functions. *Journal of Multivariate analysis*, 75 (1):62–78, 2000b.

Yijun Zuo and Robert Serfling. Structural properties and convergence results for contours of sample statistical depth functions. *Annals of Statistics*, pages 483–499, 2000c.

7 Appendix

7.1 Implementing K-Norm Mechanisms

Algorithm 5 Sampling from ℓ_1-mech

INPUT: $T(X)$, $\Delta_1(T)$, and ϵ

1: Set $m := \text{length}(T(X))$.

2: Draw $V_j \overset{\text{iid}}{\sim} \text{Laplace}((\frac{\epsilon}{\Delta_1(T)})^{-1})$ for $j = 1, \ldots, m$

3: Set $V = (V_1, \ldots, V_m)^\top$

OUTPUT: $T(X) + V$

Algorithm 6 Sampling from ℓ_2-mech (Yu et al., 2014)

INPUT: $T(X)$, $\Delta_2(T)$, and ϵ

1: Set $m := \text{length}(T(X))$.

2: Draw $Z \sim N(0, I_m)$

3: Draw $r \sim \text{Gamma}(\alpha = m, \beta = \epsilon/\Delta_2(T))$

4: Set $V = \frac{rZ}{\|Z\|_2}$

OUTPUT: $T(X) + V$

In this section, we review algorithms to implement the $\ell_1, \ell_2, \ell_\infty$-mechs. Then we give a method to implement arbitrary K-mechs. The ℓ_1-mech can be easily implemented via Algorithm 5, which only uses independent Laplace random variables. Algorithm 6, which appears in Yu et al. (2014), gives a method to sample the the ℓ_2-mech. Algorithm 7 appears in Steinke and Ullman (2017) and gives a method to sample the ℓ_∞-mech.

Algorithm 7 Sampling from ℓ_∞-mech (Steinke and Ullman, 2017)

INPUT: $T(X)$, $\Delta_2(T)$, and ϵ

1: Set $m := \text{length}(T(X))$.

2: Set $U_j \overset{\text{iid}}{\sim} U(-1, 1)$ for $j = 1, \ldots, m$

3: Draw $r \sim \text{Gamma}(\alpha = m + 1, \beta = \epsilon/\Delta_\infty(T))$

4: Set $V = r \cdot (U_1, \ldots, U_m)^\top$

OUTPUT: $T(X) + V$

In general, sampling from the K-Norm mechanisms is non-trivial. Hardt and Talwar (2010, Remark 4.2) gives a method of sampling from K-mech, provided that one can determine whether a point is in K. Precisely, we require a function $I_K : \mathbb{R}^m \to \{0, 1\}$ given by $I_K(u) = 1$ if $u \in K$ and $I_K(u) = 0$ otherwise. in Algorithm 8, we propose a procedure to sample from K using rejection sampling (see Bishop (2006, Chapter 11) for an introduction).

Example 7.1. Back to the setting of Example 2.10. The space $K_T = \text{span}(S_T)$ is

$$K_T = \left\{ (u_1, u_2) \in [-2, 2]^2 \middle| |u_2| \leq \begin{cases} 2 - (1 - u_1)^2 & \text{if } u_1 \geq 1 \\ 2 - (u_1 + 1)^2 & \text{if } u1 < -1 \end{cases} \right\}$$

Then for a vector $u = (u_1, u_2) \in \mathbb{R}^2$, our indicator function is $I_{K_T}(u) = 1$ if $u \in K_T$ and $I_{K_T}(u) = 0$ otherwise.

In Hardt and Talwar (2010), they propose a random grid-walk procedure to sample from K. However, this only gives approximate sampling. On the other hand, Algorithm 8 is easily implemented, and gives exact uniform sampling from K.

Algorithm 8 Sampling from K-Norm Mechanism with Rejection Sampling

INPUT: ϵ, $\Delta_\infty(T)$, $\Delta_K(T)$, $I_K(\cdot)$, $T(X)$

1: Set $m = \text{length}(T(X))$.
2: Draw $r \sim \text{Gamma}(\alpha = m+1, \beta = \epsilon/\Delta_K(T))$
3: Draw $U_j \overset{\text{iid}}{\sim} \text{Uniform}(-\Delta_\infty(T), \Delta_\infty(T))$ for $j = 1, \ldots, m$
4: Set $U = (U_1, \ldots, U_m)^\top$
5: If $I_K(U) = 1$, release $T(X) + U$, else go to 3).

7.2 Proofs and Technical Lemmas

Lemma 7.2. *Let $\mathscr{S}$ be a colletion of (Lebesgue) measurable sets in $\mathbb{R}^m$ such that $\arg\inf_{S\in\mathscr{S}} \lambda(S)$ is unique, where $\lambda(\cdot)$ is Lebesgue measure. Let $A : \mathbb{R}^m \to \mathbb{R}^m$ be an invertible linear transformation. Then $\arg\inf_{T\in A\mathscr{S}} \lambda(T) = A\left(\arg\inf_{S\in\mathscr{S}} \lambda(S)\right)$.*

Proof. First note that for any measurable S,

$$\lambda(AS) = \int_{AS} 1 \, dx = \int_S |\det(A^{-1})| \, du = |\det(A^{-1})|\lambda(S),$$

where we use the change of variables formula. Thus,

$$\arg\inf_{T\in A\mathscr{S}} \lambda(T) = A \arg\inf_{S\in\mathscr{S}} \lambda(AS) = A \arg\inf_{S\in\mathscr{S}} |\det(A^{-1})|\lambda(S) = A \arg\inf_{S\in\mathscr{S}} \lambda(S). \qquad \square$$

Lemma 7.3. *Let X be a random variable on $\mathbb{R}^m$ which is unimodal (center zero), continuous, and decreasing away from the center (i.e. $f_X(x) \le f_X(ax)$ for $a \in [0,1]$), and such that for all $t > 0$, $P(\{x \mid f_X(x) = t\}) = 0$. Then*

1. *S_X^α is unique,*
2. *$S_X^\alpha \subset S_X^\beta$ for $\alpha \le \beta$ (nested),*
3. *$cS_X^\alpha \subset S_X^\alpha$ for $c \in [0,1]$ (linear closure wrt center),*
4. *$S_{Ax}^\alpha = AS_X^\alpha$ for any linear, invertible map $A : \mathbb{R}^m \to \mathbb{R}^m$.*

Proof. 1. Since X is continuous and the sets $\{x \mid f_X(x) = t\}$ have probability zero for all $t > 0$, there exists $C(\alpha)$ such that $S_X^\alpha = \{x \mid f_X(x) \ge C(\alpha)\}$, where $C(\alpha)$ is a decreasing function of α (Casella and Berger, 2002, Theorem 9.3.2). This establishes uniqueness.

2. Let $x \in S_X^\alpha$. Then $f_X(x) \ge C(\alpha) \ge C(\beta)$. Hence, $x \in S_X^\beta$.

3. We calculate

$$cS_X^\alpha = \{cx \mid f_X(x) \ge C(\alpha)\} = \{y \mid f_X(c^{-1}y) \ge C(\alpha)\} \subset \{y \mid f_X(y) \ge C(\alpha)\} = S_X^\alpha,$$

where we use the fact that $f_X(a^{-1}y) \le f_X(y)$.

4. Define $\mathscr{S}_X^\alpha = \{S \mid P(X \in S) \geq \alpha\}$. Then $S_X^\alpha = \arg\inf_{S \in \mathscr{S}} \lambda(S)$, where $\lambda(\cdot)$ is Lebesgue measure. Then

   $$A\mathscr{S}_X^\alpha = \{AS \mid P(X \in S) \geq \alpha\} = \{T \mid P(X \in A^{-1}T) \geq \alpha\} = \{T \mid P(AX \in T) \geq \alpha\} = \mathscr{S}_{AX}^\alpha.$$

   Taking the infinum with respect to Lebesgue measure on both sides yields $AS_X^\alpha = S_{AX}^\alpha$. We are able to pass A in front of the infimum by Lemma 7.2. □

Proof of Lemma 3.4. We will compute the moment generating function (MGF) of $\|V\|_K$. Let $0 \leq t \leq a$. Call α the integrating constant. Then

$$\begin{aligned}
\mathbb{E}\exp(t\|V\|_K) &= \alpha^{-1} \int \cdots \int \exp(t\|v\|_K) \exp(-a\|v\|_K)\ dv_1 \ldots, dv_m \\
&= \alpha^{-1} \int \cdots \int \exp(-(a-t)\|v\|_K)\ dv_1, \ldots, dv_m \\
&= (a-t)^{-k} a^k \\
&= (1 - t/a)^{-k},
\end{aligned}$$

where we applied a u-substitution, noting that the integrand is of the same form as f_V. We identify this as the MGF of the random variable Gamma(m, a). □

Proof of Lemma 3.5. First note that S_V^α is of the form $S_V^\alpha = \{v \mid f_V(v) \geq t\}$ for some t (Casella and Berger, 2002, Theorem 9.3.2). Since $f_V(v)$ is an increasing function of $\|v\|_K$, equivalently, we have $S_V^\alpha = \{v \mid \|v\|_K \geq t\}$. We must determine the value of t such that $P(V \in \{v \mid \|v\|_K \geq t\}) = \alpha$. Recall from Lemma 3.4 that $\|V\|_K \sim$ Gamma(m, a). We conclude that t is the α quantile of Gamma(m, a). □

Proof of Theorem 3.6. By the previous lemma, we know that for all $\alpha \in (0, 1)$, $S_V^\alpha = \{x \mid \|x\|_{K_V} \leq t_\alpha\} = t_\alpha \cdot K_V$ and $S_W^\alpha = \{x \mid \|x\|_{K_W} \leq t_\alpha\} = t_\alpha \cdot K_W$ for the same value of t. Since $K_V \subset K_W$, we have that $S_V^\alpha \subset S_W^\alpha$ for all $\alpha \in (0, 1)$. □

Proof of Theorem 3.11. All we have to show is that D_X is a depth function for unimodal, continuous, decreasing random variables. For simplicity, we assume that the center of X is zero.

(A1)

$$\begin{aligned}
D_{AX}(Ax) &= 1 - \inf\{\alpha \mid Ax \in S_{Ax}^\alpha\} \\
&= 1 - \inf\{\alpha \mid x \in A^{-1} S_{Ax}^\alpha\} \\
&= 1 - \inf\{\alpha \mid X \in A^{-1} A S_X^\alpha\} \\
&= 1 - \inf\{\alpha \mid x \in S_X^\alpha\} \\
&= D_X(x),
\end{aligned}$$

where we use property 4) of Lemma 7.3.

(A2) By property 2) of Lemma 7.3, we know that the sets S_X^α are nested. So, the minimum value of D_X is attained at the mode, which has depth of 1.

(A3) Since we assume that the center is at $x_0 = 0$, it suffices to show that $D_X(x) \leq D_X(ax)$ for $a \in [0, 1]$. Let $y \in \mathbb{R}^m$. First we will show that $\{\alpha \mid y \in aS_X^\alpha\} \subset \{\alpha \mid y \in S_X^\alpha\}$. Let $\alpha \in \{\alpha \mid y \in aS_X^\alpha\}$. Then $y \in aS_X^\alpha \subset S_X^\alpha$, by property 3) of Lemma 7.3. So, $\alpha \in \{\alpha \mid y \in S_X^\alpha\}$. It follows that $\inf\{\alpha \mid y \in aS_X^\alpha\} \geq \inf\{\alpha \mid y \in S_X^\alpha\}$. Finally, by choosing $y = ax$ we have that

$$D_X(x) = D_{aX}(y) = 1 - \inf\{\alpha \mid y \in aS_X^\alpha\} \leq 1 - \inf\{\alpha \mid y \in S_X^\alpha\} = D_X(y) = D_X(ax),$$

where we use property (A1) of Definition 3.8, and property 4) of Lemma 7.3.

(A4) We will show that $\lim_{t\to\infty} D_X(tx) = 0$ for all $x \in \mathbb{R}^m$. Let $\gamma > 0$ and $x \in \mathbb{R}^m$ be given. By assumption, we have that for all x, $\lim_{t\to\infty} f_X(tx) = 0$. So, there exists $t > 0$ such that $f(tx) \leq C(1 - \gamma)$, where $C(\cdot)$ is the function defined in the proof of Lemma 7.3 part 1. Then $D_X(tx) = 1 - \inf\{\alpha \mid f_X(tx) \geq C(\alpha)\} \leq 1 - (1 - \gamma) = \gamma$, where we use the fact that $1 - \gamma \leq \inf\{\alpha \mid f_X(tx) \geq C(\alpha)\}$.

□

Proof of Proposition 3.14. We can write the integrand as $-\log f(V) = -\log\left((\epsilon/\Delta)^m \frac{1}{m!\lambda(K)}\right) + \frac{\epsilon}{\Delta}\|V\|_K$. Then

$$\begin{aligned} H(V) &= \mathbb{E} - \log f(V) \\ &= \log\left((\Delta/\epsilon)^m m!\lambda(K)\right) + \frac{\epsilon}{\Delta}\mathbb{E}\|V\|_K \\ &= \log\left((\Delta/\epsilon)^m m!\lambda(K)\right) + \frac{\epsilon}{\Delta}\frac{m\Delta}{\epsilon}, \end{aligned}$$

where we recall from Lemma 3.4 that $\|V\|_K \sim \text{Gamma}(m, \epsilon/\Delta)$. □

Proof of Lemma 3.16. 1. Recall from Hardt and Talwar (2010, Remark 4.2) that $V \stackrel{d}{=} R \cdot U$, where $R \sim \text{Gamma}(m + 1, a)$, $U \sim \text{Unif}(K)$, and $R \perp\!\!\!\perp U$. Then $\|V\|_K = R\|U\|$, and $\frac{V}{\|V\|_K} = \frac{RU}{R\|U\|_K} = \frac{U}{\|U\|_K}$. So, it suffices to show that $U \perp\!\!\!\perp \frac{U}{\|U\|_K}$. To this end, we will derive the conditional cdf of $\|U\|_K$ given that $\frac{U}{\|U\|_K} = e$ for some $\|e\|_K = 1$.

First, for any $\gamma > 0$, we define the set $\text{Cone}(e, \gamma) = \left\{v \in \mathbb{R}^m \mid \arccos\left(\frac{v^\top e}{\|v\|_2 \cdot \|e\|_2}\right) \leq \gamma\right\}$, which is the set of all vectors v whose angle from e is less than γ. Then for any $t \geq 0$, we can express the conditional cdf as

$$\begin{aligned} P\left(\|U\|_K \leq t \;\middle|\; \frac{U}{\|U\|_K} = e\right) &= \lim_{\gamma\to 0} \frac{P(U \in (tK \cap \text{Cone}(e, \gamma)))}{P(U \in (K \cap \text{Cone}(e, \gamma)))} \\ &= \lim_{\gamma\to 0} \frac{\lambda(tK \cap \text{Cone}(e, \gamma))/\lambda(K)}{\lambda(K \cap \text{Cone}(e, \gamma))/\lambda(K)} \\ &= \lim_{\gamma\to 0} \frac{t^m \lambda(K \cap \text{Cone}(e, \gamma))}{\lambda(K \cap \text{Cone}(e, \gamma))} \\ &= t^m. \end{aligned}$$

We see that the conditional cdf $P(\|U\|_K \le t \mid \frac{U}{\|U\|_K} = e)$ does not depend on e. We conclude that $\|U\|_K \perp\!\!\!\perp \frac{U}{\|U\|_K}$ and hence $\|V\|_K \perp\!\!\!\perp \frac{V}{\|V\|_K}$.

2. Since $V \in \text{span}(e)$, we know that $|V^\top e| = \|V\|_2 \cdot \|e\|_2 = \|V\|_2$. Then

$$|V^\top e| = \|V\|_2 = \left\| \|V\|_K \cdot \frac{V}{\|V\|_K} \right\|_2 = \|V\|_K \left\| \frac{e}{\|e\|_K} \right\|_2 = \frac{\|V\|_K}{\|e\|_K} \sim \text{Gamma}(m, a\|e\|_K),$$

where we use the fact that $\frac{V}{\|V\|_K} = \pm \frac{e}{\|e\|_K}$, that $\|V\|_K$ is independent of $\frac{V}{\|V\|_K}$, and that $\|V\|_K \sim \text{Gamma}(m, a)$ from Lemma 3.4. □

Proof of Theorem 3.17. By Lemma 3.16, $W = \left(V_K^\top e \middle| V_K \in E\right)$ is distributed as $\text{Gamma}(m, \frac{\epsilon\|e\|_K}{\Delta_K})$, which has variance $\frac{m\Delta_K}{\epsilon^2\|e\|_K}$. Minimizing the variance between the K-norm and H-norm is equivalent to maximizing $\frac{\|e\|_K}{\Delta_K}$. Note that $\frac{\|e\|_K}{\Delta_K}$ is the same value as the diameter of the set $\text{span}(e) \cap (\Delta_K \cdot K)$. Since $\Delta_K \cdot K \subset \Delta_H \cdot H$, we have that $\text{span}(e) \cap (\Delta_K \cdot K) \subset \text{span}(e) \cap (\Delta_H \cdot H)$. The result follows. □

Proof of Theorem 4.1. First we will assume that $r(\theta)$ is twice-differentiable and $\Theta = \mathbb{R}^m$. At the end of the proof we will use the results in Kifer et al. (2012) to extend the proof to arbitrary convex $r(\theta)$ and arbitrary convex sets Θ.

By Awan et al. (2019, Proposition 2.3), it suffices to show that for all $a \in \mathbb{R}^m$ and all $X, X' \in \mathscr{X}^n$ with $\delta(X, X') = 1$ we have

$$\frac{\text{pdf}(\theta_{DP} = a \mid X)}{\text{pdf}(\theta_{DP} = a \mid X')} \le \exp(\epsilon).$$

To this end, let $a \in \mathbb{R}^m$ and X, X' be such that $\delta(X, X') = 1$. Then if $\theta_{DP} = a$, we have $a = \arg\min_{\theta \in \mathbb{R}^m} n\hat{\mathscr{L}}(a; X) + r(\theta) + \frac{\gamma}{2}\theta^\top\theta + V^\top\theta$. By taking the gradient with respect to θ and setting it equal to zero, we can solve for V as a function of a: $V(a; X) = -(n\nabla\hat{\mathscr{L}}(a; X) + \nabla r(a) + \gamma a)$. Then applying this one-to-one change of variables, we get

$$\frac{\text{pdf}(\theta_{DP} = a \mid X)}{\text{pdf}(\theta_{DP} = a \mid X')} = \frac{f(V(a; X) \mid X)}{f(V(a; X') \mid X')} \frac{|\det \nabla V(a; X')|}{|\det \nabla V(a; X)|}$$

We will bound these two factors separately. First, we have $\frac{f(V;X)}{f(V;X')} \le \exp(\epsilon q)$, by Proposition 2.8. As $\delta(X, X') = 1$, without loss of generality, assume that $X_i = X_i'$ for all $i = 1, \ldots, n-1$. Call $A = \nabla V(a; X)$, $B = \nabla V(a; X')$, and $C = \sum_{i=1}^{n-1} \nabla^2\ell(a; X_i) + \nabla^2 r(a) + \gamma I_m$, where I_m is the $m \times m$ identity matrix. Note that $A = C + \nabla^2\ell(a; X_n)$ and $B = C + \nabla^2\ell(a; X_n')$. Then

we have

$$\left|\frac{\det \nabla V(a;X')}{\det \nabla V(a;X)}\right| = \left|\frac{\det(B)}{\det(A)}\right| = \left|\frac{\det(C+\nabla^2\ell(a;X'_n))}{\det(C+\nabla^2\ell(a;X_n))}\right| \tag{5}$$

$$= \left|\frac{\det(C)\det(I_m + C^{-1}\nabla^2\ell(a;X'_n))}{\det(C)\det(I_m + C^{-1}\nabla^2\ell(a;X_n))}\right| \tag{6}$$

$$\leq \frac{1+\frac{\lambda}{\gamma}}{|\det(I_m + C^{-1}\nabla^2(a;X_n))|} \tag{7}$$

$$\leq 1 + \frac{\lambda}{\gamma} = \exp(\epsilon(q-1)) \tag{8}$$

To justify the inequality in (7), note that $C^{-1}\nabla^2\ell(a;X'_n)$ is positive definite of rank at most 1. We know that $\nabla^2\ell(a;X'_n)$ has at most one nonzero eigenvalue. Furthermore, γ is a lower bound on the eigenvalues of C,since C is the sum of positive definite functions and γI_m. So, $C^{-1}\nabla^2\ell(a;X'_n)$ has at most one nonzero eigenvalue, which is bounded between 0 and λ/γ.

Next we justify the inequality in (8). Since $C^{-1}\nabla^2\ell(a;X_n)$ is positive definite, all of its eigenvalues are non-negative. Thus, all eigenvalues of $(I + C^{-1}\nabla^2\ell(a;X_n))$ are greater than or equal to 1. Hence, $\det(I + C^{-1}\nabla^2\ell(a;X_n)) \geq 1$.

The last equality just uses the fact that $\gamma = \frac{\lambda}{\exp(\epsilon(q-1))-1}$. Finally, we combine our bounds:

$$\frac{\text{pdf}(\theta_{DP}=a;X)}{\text{pdf}(\theta_{DP}=a;X')} = \frac{f(V;X)}{f(V;X')}\left|\frac{\det(\nabla V(a;X'))}{\det(\nabla V(a;X))}\right| \leq \exp(\epsilon q)\exp(\epsilon(q-1)) = \exp(\epsilon).$$

Theorem 1 in Kifer et al. (2012) on successive approximations extends without modification to Definition 2.2. Using this theorem along with the techniques detailed in Appendix C.2, C.3 of Kifer et al. (2012), we extend this proof to arbitrary convex functions $r(\theta)$ and arbitrary convex sets Θ. □