General Staircase Mechanisms for Optimal Differential Privacy

Alex Kulesza
Google Research

Ananda Theertha Suresh
Google Research

Yuyan Wang
Google Research

Abstract

We derive the optimal differentially private additive noise mechanism for queries in $\mathbb{R}^d$ when sensitivity and error are defined by an arbitrary norm $\|\cdot\|_K$. The optimal mechanism is a generalization of the staircase mechanism, which is known to be optimal under the ℓ_1 norm when $d \leq 2$; we extend the mechanism and its guarantee to arbitrary norms and dimensions, proving a conjecture of Geng et al. [2015] along the way. The generalized staircase mechanism we derive can be viewed as a refinement of the K-norm mechanism of Hardt and Talwar [2010], with improvements particularly evident in the low-privacy regime as $\varepsilon \to \infty$. We show how to implement the generalized staircase mechanism efficiently, given an efficient algorithm for sampling the unit K-norm ball, and demonstrate that it significantly reduces error in realistic settings, including under non-standard norms that are common in practice, and across a range of error metrics.

1 INTRODUCTION

Differential privacy has been applied to many statistical estimation tasks, including, for instance, mean estimation [Bun and Steinke, 2019], median estimation [Asi and Duchi, 2020], histogram estimation [Xu et al., 2013], and heavy hitter estimation [Bassily et al., 2017]. In many such tasks, the output value can be high-dimensional, and a standard approach is to compute the non-private metric and add Laplace noise or Gaussian noise. However, the resulting mechanism is generally suboptimal.

This was first observed by Hardt and Talwar [2010], who proposed private additive noise mechanisms that are specific to the *sensitivity space* of the query, defined as the set of all possible differences in the query value between neighboring datasets (see Definition 2). Their algorithm, the K-norm mechanism, generalizes the Laplace mechanism and often performs better in practice. Hardt and Talwar [2010] gave bounds on the performance of their algorithm, and showed that it can be implemented in polynomial time and has better utility compared to Laplace noise for the task of answering multiple linear queries. More recently, Joseph and Yu [2024] proposed algorithms to efficiently sample the K-norm mechanism for sum, count, and majority queries.

Separately, Geng and Viswanath [2014] initiated a line of work on optimal mechanisms for low-dimensional queries called *staircase* mechanisms, so named because they add noise with a density that decays in piecewise constant steps, unlike smooth Laplace or Gaussian noise. Geng and Viswanath [2014] proposed the one-dimensional staircase mechanism and demonstrated both theoretically and empirically that it outperforms the Laplace mechanism. Subsequently, Geng et al. [2015] generalized this approach to d-dimensional staircase mechanisms and showed that for $d \leq 2$ they outperform the Laplace mechanism (and all other additive noise mechanisms) under ℓ_1 sensitivity. Geng et al. [2020] later extended the one-dimensional staircase mechanism to (ε, δ)-differential privacy; more recently, Kulesza et al. [2023] proposed a staircase-like mechanism for a sub-problem arising in one-dimensional mean estimation.

In this paper, we show how these lines of work can be brought together, establishing that a generalized staircase mechanism is the optimal additive noise mechanism whenever sensitivity and error are determined by the same norm K, even in high dimension. We show that this mechanism can be implemented efficiently in a manner similar to the K-norm mechanism, assuming the existence of an efficient sampler for the K-norm unit ball. Moreover, we demonstrate empirically that the improvements over the Laplace and K-norm mechanisms are significant in realistic settings, and that the

Proceedings of the 28th International Conference on Artificial Intelligence and Statistics (AISTATS) 2025, Mai Khao, Thailand. PMLR: Volume 258. Copyright 2025 by the author(s).

generalized staircase mechanism still performs better even when error is measured using a metric other than the K-norm[1].

2 PRELIMINARIES

Let $\mathcal{D}$ denote the set of all possible input datasets. Differentially private mechanisms are then defined as follows.

Definition 1 (Differential privacy [Dwork et al., 2014])**.** *A mechanism $A : \mathcal{D} \to \mathbb{R}^d$ satisfies ε-differential privacy if for any two neighboring datasets D, D' and for any output $\mathcal{S} \subseteq \mathbb{R}^d$, it holds that*

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

Two popular notions of neighboring datasets are used. The first, called the *add-remove* model [Dwork et al., 2014, Definition 2.4], defines D and D' as neighboring if and only if $|D \setminus D'| + |D' \setminus D| = 1$. The second, called the *swap* model [Dwork et al., 2006, Vadhan, 2017], defines two datasets D and D' as neighboring if and only if $|D \setminus D'| = 1$ and $|D' \setminus D| = 1$. In this paper, we are ambivalent about the semantics of the neighboring relation; instead, we use the notion of a *sensitivity space* to characterize how the target query changes on neighboring datasets.

More formally, let $q : \mathcal{D} \to \mathbb{R}^d$ denote a query that maps a dataset to a d-dimensional value.

Definition 2. *The* sensitivity space $K \subseteq \mathbb{R}^d$ *for q is given by* $K = \{x : \exists \text{ neighboring } D, D' \text{ s.t. } x = q(D) - q(D')\}$.

Notice that K is symmetric around the origin ($K = -K$), since for all neighboring pairs D, D' both $q(D) - q(D')$ and $q(D') - q(D)$ must belong to K.

In practice, sensitivity spaces are often also convex—for instance, the sensitivity space will be convex if the query is a sum of vectors from a convex set like the ℓ_1 or ℓ_2 ball—and we will assume here that K is convex. We can then define the Minkowski K-norm $\|x\|_K = \inf\{r : x \in rK\}$, where $rK = \{r \cdot z : z \in K\}$. (We note in passing that non-convex sensitivity spaces also sometimes arise in practice—see Kulesza et al. [2023] for one example—and are of interest for future work.) See Figure 1 for an illustration of a sensitivity space K.

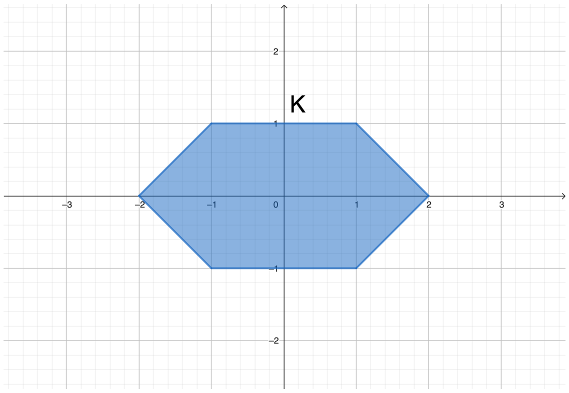

Figure 1: A sensitivity space K. Note that K is convex and symmetric about the origin. See Figure 2 for an example of a staircase distribution for this sensitivity space.

Our aim is to find the optimal mechanism for q among all *additive noise mechanisms*, which produce a differentially private estimate of $q(D)$ by adding noise sampled from a suitable distribution $\mathcal{P}$; i.e.,

$$A(D) = q(D) + Z \ ,$$

where $Z \sim \mathcal{P}$. Common noise distributions include the Laplace distribution and the Gaussian distribution.

Definition 3. *We call a noise distribution $\mathcal{P}$ (K, ε)-*differentially private *if, for all $S \subseteq \mathbb{R}^d$ and all $x \in K$,*

$$\mathcal{P}(S) \leq e^{\varepsilon} \mathcal{P}(S + x) \ ,$$

where $S + x = \{y + x : y \in S\}$.

Note that the additive mechanism A defined above is ε-differentially private whenever $\mathcal{P}$ is a (K, ε)-differentially private noise distribution.

Definition 4. *The K-norm mechanism of Hardt and Talwar [2010] is an additive noise mechanism using the (K, ε)-differentially private noise density*

$$p(x) \propto e^{-\varepsilon \|x\|_K} \ .$$

2.1 Optimal Additive Noise

Let $\ell : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ denote a loss function that measures the quality of the output with respect to the true value of q, and let

$$\mathcal{L}_D(A, q) = \mathbb{E}\left[\ell(A(D), q(D))\right]$$

be the expected loss of mechanism A on dataset D.

For our theoretical results, we will assume the loss function is given by the K-norm:

$$\ell(A(D), q(D)) = \|A(D) - q(D)\|_K \ .$$

Hence, for the additive mechanism A defined above,

$$\mathcal{L}_D(A, q) = \mathbb{E}\left[\ell(A(D), q(D))\right] = \mathbb{E}_{Z \in \mathcal{P}}\left[\|Z\|_K\right] ,$$

[1] The code can be found at `https://github.com/google-research/google-research/tree/master/general_staircase_mechanism`

independent of the underlying dataset D.

The goal of this work is to identify the noise distribution Z that minimizes $\mathbb{E}_{Z \in \mathcal{P}}[\|Z\|_K]$, subject to the corresponding additive mechanism A remaining differentially private.

Definition 5. *We call a noise distribution $\mathcal{P}^*$ (K, ε)-optimal if it is (K, ε)-differentially private and, for all (K, ε)-differentially private noise distributions $\mathcal{P}$,*

$$E_{Z \sim \mathcal{P}^*}[\|Z\|_K] \leq E_{Z \sim \mathcal{P}}[\|Z\|_K].$$

When K is the ℓ_1 ball and $d \leq 2$, Geng et al. [2015] showed that the staircase mechanism is the optimal additive noise mechanism. However, they only conjectured the same result for $d > 2$. We prove their conjecture and extend the result to general K.

3 THEORETICAL RESULTS

3.1 Optimal Additive Noise Mechanisms

To identify the optimal additive noise mechanism, we first narrow down the list of candidates, showing that there exists an optimal mechanism that places constant probability density on noise values having the same K-norm.

Theorem 1 (Optimality of equicontour mechanisms)**.** *For any sensitivity set K and privacy parameter $\varepsilon > 0$, there exists a (K, ε)-optimal noise distribution whose density p takes the form $p(x) = f(\|x\|_K)$.*

To prove Theorem 1, we show that any (K, ε)-differentially private noise mechanism (including any optimal mechanism) can be "averaged out" across the K-spheres without reducing privacy or increasing loss. See Section 6 for more details.

We then proceed to examine the class of equicontour mechanisms in more detail. Many mechanisms, including the K-norm mechanism of Hardt and Talwar [2010], are equicontour, but we show that the optimal one is a generalization of the staircase mechanism.

Definition 6 (Generalized staircase mechanism)**.** *A (K, ε)-differentially private noise distribution $\mathcal{P}$ on $\mathbb{R}^d$ is a* generalized staircase distribution *(and its corresponding additive mechanism is a* generalized staircase mechanism*) if it has a density $p(x) = f_\gamma(\|x\|_K)$ for some $\gamma \in (0, 1]$, where*

$$f_\gamma(r) \propto e^{-\lfloor r - \gamma \rfloor \varepsilon}.$$

Notice that f_γ is constant on the intervals $[0, \gamma)$, $[\gamma, \gamma + 1)$, $[\gamma + 1, \gamma + 2)$, etc., and decreases by a factor of e^ε on each subsequent interval. Figure 2 depicts the generalized staircase mechanism for the two-dimensional sensitivity space from Figure 1 when $\gamma = 0.6$.

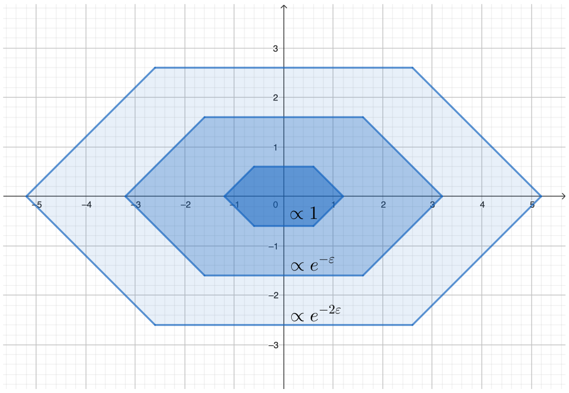

Figure 2: The first few stairs of the staircase distribution for K from Figure 1 and $\gamma = 0.6$. The text in each shaded region shows the density for that region.

Intuitively, the optimal noise density should be decreasing roughly exponentially in $\|\cdot\|_K$ to minimize error. However, if a density decays smoothly with $\|\cdot\|_K$, then we can improve its utility by moving a small amount of mass from regions of higher magnitude noise to lower magnitude noise. Because nearby density ratios are limited by privacy constraints, in the limit such improvements lead to a "staircase" shape, where as much mass as possible has been pushed into the low magnitude regions. Formally, we have the following result.

Theorem 2 (Optimality of generalized staircase mechanisms)**.** *For any sensitivity set K and privacy parameter $\varepsilon > 0$, there is a generalized staircase distribution that is (K, ε)-optimal.*

The optimal γ that maximizes the utility can be found by a grid search. However, our analysis also yields the following observation.

Lemma 1. *The optimal value of γ is independent of the sensitivity space and depends only on the privacy parameter ε and the dimension d.*

As a corollary, we show that the optimal error (measured as the expected K-norm) is also independent of the sensitivity space.

Corollary 1. *The error of the optimal generalized staircase mechanism is independent of the underlying sensitivity space K and depends only on the privacy parameter ε and the dimension d.*

In Section 6 we return to these results and provide an overview of the techniques we use to prove them. Full proofs are provided in Appendix A.

3.2 Comparison With the K-norm Mechanism

In the one-dimensional setting, Geng and Viswanath [2014] compared the staircase mechanism and the Laplace mechanism (equivalent to the K-norm mechanism in one dimension) and showed that their errors converge in the high-privacy regime where $\varepsilon \to 0$, but the error of the staircase mechanism improves exponentially over the Laplace mechanism as $\varepsilon \to \infty$. Here, we extend that analysis to the d-dimensional case.

Since both d and ε affect the error, we consider each independently. Let $r(\varepsilon, d)$ denote the ratio between the error of the generalized staircase mechanism and the error of the K-norm mechanism given the specified parameters. (This ratio is always upper-bounded by 1 since the generalized staircase mechanism is optimal.)

Theorem 3. *For a fixed ε, as $d \to \infty$, the ratio $r(\varepsilon, d)$ is monotone non-decreasing.*

Theorem 4. *For a fixed d, in the high-privacy regime as $\varepsilon \to 0$, $r(\varepsilon, d) \to 1$, hence the generalized staircase mechanism gives the same error as the K-norm mechanism. In the low-privacy regime as $\varepsilon \to \infty$, $r(\varepsilon, d) = O(exp(-\Omega(\varepsilon))$, and the generalized staircase mechanism becomes exponentially better than the K-norm mechanism.*

Proofs and a more detailed discussion can be found in Appendix A.

4 SAMPLING ALGORITHM

We next propose an algorithm to implement the generalized staircase mechanism. Our algorithm involves sampling from the discrete gamma distribution, which is defined as follows.

Definition 7 (Discrete gamma distribution). *Given parameters γ, ε, and dimension d, define $D\Gamma_{\gamma,\varepsilon,d}$ to be the distribution supported on $\{i + \gamma\}_{i=0,1,2,\ldots}$ with*

$$D\Gamma_{\gamma,\varepsilon,d}(i+\gamma) = \frac{1}{C_{\varepsilon,d}(\gamma)}(i+\gamma)^d e^{-\varepsilon i} \ ,$$

where $C_{\varepsilon,d}(\gamma)$ is the normalizer given by

$$C_{\varepsilon,d}(\gamma) = \sum_{i=0}^{\infty}(i+\gamma)^d e^{-\varepsilon i}.$$

Algorithm 1 provides an implementation for sampling from the discrete gamma distribution.

Lemma 2. *For any ε, γ, and d, Algorithm 1 returns a sample from the discrete gamma distribution $D\Gamma_{\gamma,\varepsilon,d}$, and the expected number of samples required from the Laplace distribution is at most $O(d)$.*

Algorithm 1 Sampling from the discrete gamma distribution.

```
Input: $\varepsilon$, $\gamma$, $d$.
while True do
    Sample $d$ random variables $X_1, \ldots, X_d$, i.i.d. from the  exponential distribution : $f(x) \propto e^{-\varepsilon x}$ for $x \geq 0$.
    Let their sum be $s = \sum_{i=1}^{d} X_i$.
    Let $t = \lceil s - \gamma \rceil + \gamma - s$ .
    Sample another random variable $Y$ from the Laplace $f(x) \propto e^{-\varepsilon x}$.
    If $Y \geq t$: return $\lceil s - \gamma \rceil + \lfloor Y - t \rfloor$.
end while
```

Algorithm 2 returns a sample from the generalized staircase mechanism. It is analogous to the sampling algorithm for the K-norm mechanism [Hardt and Talwar, 2010], but uses Algorithm 1 in place of sampling from the (continuous) Gamma distribution. It also relies on a subroutine that returns a uniform sample from K (under the Lebesgue measure). Such subroutines have previously been developed for a number of sensitivity spaces encountered in practice [Hardt and Talwar, 2010, Joseph and Yu, 2024].

Algorithm 2 Sampling from the generalized staircase distribution.

```
Input: Sensitivity space $K \subseteq \mathbb{R}^d$, uniform sampling subroutine $\mathcal{U}$, $\varepsilon$, $\gamma$.
Obtain a uniform sample from $K$: $x \sim \mathcal{U}(K)$.
Obtain a scalar sample from the discrete gamma distribution: $y \sim D\Gamma_{\gamma,\varepsilon,d}$.
Output $y \cdot x$.
```

It is known that the (smooth) K-norm mechanism can be sampled in a similar fashion, first uniformly drawing a sample from the sensitivity space and then rescaling it by a sample from the (continuous) gamma distribution. Here we use essentially the same method, but draw the scale from the discrete gamma distribution.

Theorem 5. *Algorithm 2 returns a sample from the generalized staircase mechanism, requiring one uniform sample from K and an expected $O(d)$ samples from the exponential distribution.*

5 EXPERIMENTS

In this section we evaluate the performance of the generalized staircase mechanism and compare it to the K-norm mechanism and the Laplace mechanism. In particular, we make the following three observations.

- **Validation of theory:** Our theoretical results show that the generalized staircase mechanism performs better than the K-norm mechanism (and every other additive mechanism, including the Laplace mechanism) when the loss of algorithm is measured by the K-norm. We demonstrate this phenomenon empirically and show that the gap increases as ε increases. In many practical settings, for moderate ε, the improvement is significant.

- **Robustness to γ:** As described in Section 3, the offset γ for the generalized staircase mechanism can be found by a grid search. We show that the utility of the mechanism is robust to the choice of γ, implying that even an imperfect search is likely to yield improved results compared to the K-norm mechanism.

- **Utility in other metrics:** Although our theoretical results show that the generalized staircase mechanism is optimal when utility is measured using the K-norm, practitioners may use other metrics. We show that the generalized staircase mechanism's improved performance extends to a variety of different error metrics, providing benefits over existing mechanisms even in settings where it may not be theoretically optimal.

Parameters of interest. Since we are using additive noise mechanisms, the final utility is independent of the underlying dataset and depends only on the sensitivity space K, the privacy parameter ε, the underlying dimension d, and the metric used for measuring performance.

5.1 Validation Of Theory

We first demonstrate that when performance is measured using the matching norm $\|\cdot\|_K$, the generalized staircase mechanism outperforms the K-norm mechanism (which, in turn, matches or outperforms the Laplace mechanism [Hardt and Talwar, 2010]). Figure 3 compares the staircase and K-norm mechanisms for $d = 3$ as ε varies from 2^{-2} to 2^5 using the optimal value of γ. Observe that the staircase mechanism significantly outperforms the K-norm mechanism for roughly $\varepsilon \geq 2$. (Note that, following Corollary 1, the choice of K does not affect these results.) On the other hand, Figure 4 compares the two mechanisms for fixed $\varepsilon = 4$ over varying dimension $d \in \{1, \ldots, 8\}$, again using the optimal γ. As d increases, the advantage of the generalized staircase mechanism decreases; however, note that for larger ε the improvement may still be significant for larger d. For additional discussion on the effects of d, ε and γ, see Appendix B.

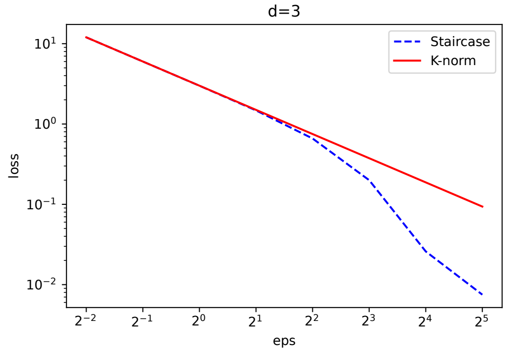

Figure 3: Expected error of the staircase/K-norm mechanisms for varying ε, $d = 3$, and optimal γ.

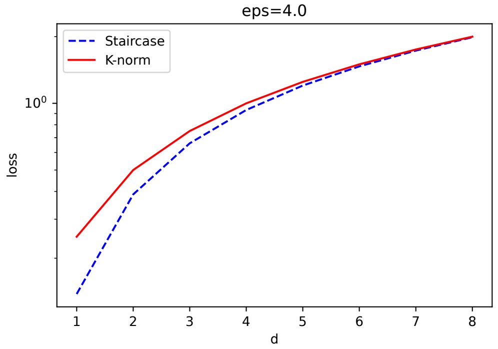

Figure 4: Expected error of the staircase/K-norm mechanisms for varying d, $\varepsilon = 4$, and optimal γ.

5.2 Robustness To γ

Figure 5 shows the error of the generalized staircase mechanism as a function of γ, with the K-norm mechanism error as a reference. Although the utility varies significantly overall, for a wide range of γ around the optimal γ^* the staircase mechanism outperforms the K-norm mechanism. Further, around γ^* the expected error changes slowly. Thus, despite the difficulty of analytically finding γ^*, we might still get significantly improved performance choosing γ using simple empirical techniques such as grid search.

5.3 Utility In Other Metrics

We next demonstrate the performance of the generalized staircase mechanism across different combinations of sensitivity spaces and error metrics using ℓ_p norms for $p = 1, 2, \infty$. These norms include many realistic sensitivity spaces—for example, the unit ℓ_∞ ball is the sensitivity space for a sum over a dataset of vectors where each entry is in $[-1, 1]$—as well as commonly used error metrics. Figure 6 compares the mechanisms for each combination of sensitivity space

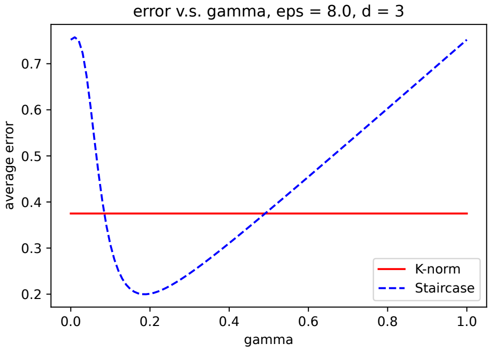

Figure 5: Expected error of the staircase/K-norm mechanisms for varying γ, $\varepsilon = 8.0$, and $d = 3$.

and error norm. γ remains fixed to the optimal value for the sensitivity norm, even when evaluating under a different norm. Observe that, in all cases, the generalized staircase mechanism significantly outperforms the K-norm mechanism, even when the error metric does not match.

5.3.1 A Non-standard Sensitivity Norm

In order to demonstrate the generality of the proposed mechanism, we adopt the "sum" setting studied by Joseph and Yu [2024]. In this setting, the dataset contains records from many users, each a vector in $[-1,1]^d$, and the target query sums all of the records. However, each user's contributed vector is also restricted to at most k nonzero components. (This is motivated by real-world systems that specify both ℓ_0 and ℓ_∞ bounds [Amin et al., 2022, Rogers et al., 2020].)

Joseph and Yu [2024] show that the convex hull of the sensitivity space for this setting is the intersection of the $[-1,1]^d$ hypercube and the ℓ_1 ball $\{v : \|v\|_1 \leq k\}$. For example, when $(d,k) = (3,2)$, the result is as shown in Figure 7 (for each of the eight orthants). This somewhat unusual sensitivity space gives rise to a non-standard K-norm, but we can still apply the generalized staircase mechanism.

Figure 8 shows the results when measuring error using the standard ℓ_p norms from above. The generalized staircase mechanism dramatically outperforms the order-optimal K-norm mechanism and the frequently-used Laplace mechanism across all error metrics, even though none of them matches the sensitivity space.

6 ANALYSIS OUTLINE

Finally, we return to the theoretical results from Section 3 and show some of the techniques used to prove them. The remaining details can be found in Appendix A.

6.1 Optimality Of Equicontour Mechanisms

We start by proving that ε-DP additive noise distributions have densities.

Lemma 3. *Let $\mathcal{P}$ be a (K,ε)-differentially private noise distribution. Then $\mathcal{P}$ has a density (with respect to the Lebesgue measure λ). That is, there exists a function $p : \mathbb{R}^d \to [0,\infty)$ such that, for any Lebesgue-measurable $Z \subseteq \mathbb{R}^d$,*

$$\mathcal{P}(Z) = \int_Z p(z)\, d\lambda(z) \ .$$

Proof. By the Radon-Nikodym theorem, it suffices to show that $\mathcal{P}$ is absolutely continuous with respect to the Lebesgue measure, i.e., that $\lambda(Z) = 0$ implies $\mathcal{P}(Z) = 0$. Suppose to the contrary that there exists a Z with $\lambda(Z) = 0$ but $\mathcal{P}(Z) > 0$. For any $x \in K$, we have $\mathcal{P}(Z + x) \geq e^{-\varepsilon}\mathcal{P}(Z)$. Now consider the expected $\mathcal{P}$-measure of the intersection of Z and $Z + x$ when x is drawn uniformly at random:

$$\begin{aligned}
&E_{x\in K}\left[\mathcal{P}(Z \cap (Z+x))\right] \\
&= \int_K \mathcal{P}(Z \cap (Z+x))\, d\lambda(x) \\
&= \int_K \int \mathbf{1}(z \in Z)\mathbf{1}(z \in Z + x)\, d\mathcal{P}(z)\, d\lambda(x) \\
&= \int \int_K \mathbf{1}(z \in Z)\mathbf{1}(z \in Z + x)\, d\lambda(x)\, d\mathcal{P}(z) \\
&= \int \mathbf{1}(z \in Z) \left[\int_K \mathbf{1}(x \in z - Z)\, d\lambda(x)\right] d\mathcal{P}(z) \\
&= 0 \ ,
\end{aligned}$$

where we use Tonelli's theorem to switch the order of integration and then the fact that $\lambda(z - Z) = 0$ (since $\lambda(Z) = 0$).

Thus, there exists $x \in K$ such that $\mathcal{P}(Z \cap (Z + x)) = 0$. Let $Z_1 = Z \cup (Z + x)$; then

$$\begin{aligned}
\mathcal{P}(Z_1) &= \mathcal{P}(Z) + \mathcal{P}(Z + x) - \mathcal{P}(Z \cap (Z + x)) \\
&\geq (1 + e^{-\varepsilon})P(Z) \ .
\end{aligned}$$

Since $\lambda(Z_1)$ is still zero, we can iterate this process, obtaining a sequence $Z_1, Z_2, \ldots$ with $\mathcal{P}(Z_i) \geq (1 + e^{-\varepsilon})^i\mathcal{P}(Z)$. For sufficiently large i, $\mathcal{P}(Z_i) > 1$, which contradicts the fact that $\mathcal{P}$ is a probability distribution. Thus the set Z cannot exist. □

Next, we show that there exists a density for $\mathcal{P}$ that satisfies the privacy constraints pointwise over all of $\mathbb{R}^d$.

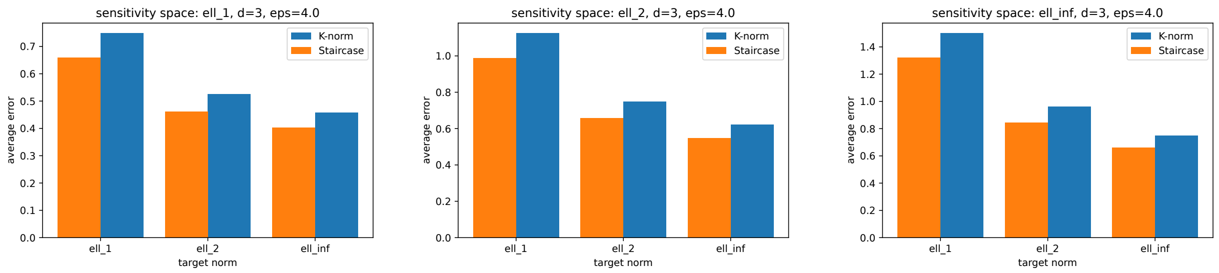

Figure 6: Expected error of the staircase/K-norm mechanisms for different combinations of sensitivity space and target error norm with $\varepsilon = 4$, $d = 3$, and optimal γ (for the sensitivity norm).

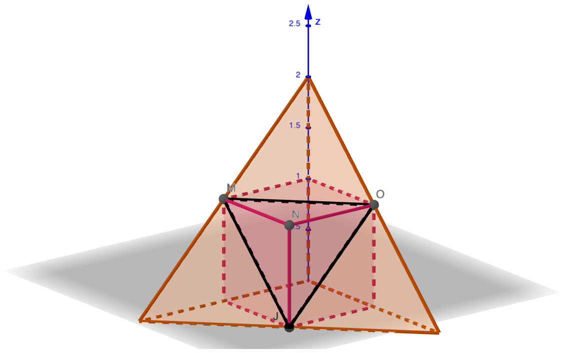

Figure 7: The "sum" polytope in the first orthant for $d = 3$ and $k = 2$. It is equivalent to the unit cube with the pyramid (M, N, O, J) removed.

Lemma 4. *Let $\mathcal{P}$ be a (K, ε)-differentially private noise distribution. Then $\mathcal{P}$ has a finite density p such that, for all $x, y \in \mathbb{R}^d$, $x - y \in K$,*

$$p(x) \leq e^{\varepsilon} p(y) \ .$$

Proof. Let $B(x, r)$ denote the ball of radius r centered at $x \in \mathbb{R}^d$, and define

$$p(x) = \limsup_{r \to 0^+} \frac{\mathcal{P}(B(x,r))}{|B(x,r)|} \ .$$

Let p_0 be an arbitrary density of $\mathcal{P}$. By the Lebesgue differentiation theorem,

$$p_0(x) = \lim_{r \to 0^+} \frac{\mathcal{P}(B(x,r))}{|B(x,r)|}$$

for almost all x, and thus $p = p_0$ almost everywhere, since the limit superior agrees with the limit whenever it exists. Thus p is also a density of $\mathcal{P}$. (In particular, it is the *upper density* of $\mathcal{P}$.) It remains to show that p satisfies the pointwise privacy constraint and is finite.

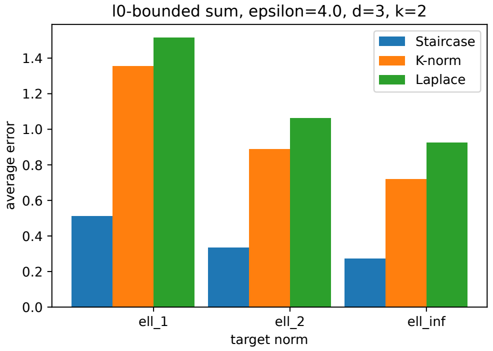

Figure 8: Expected error of the staircase, K-norm, and Laplace mechanisms for the "sum" sensitivity space from Joseph and Yu [2024] under various error norms with $\varepsilon = 4$, $d = 3$, $k = 2$, and optimal γ (for the sensitivity norm).

When $x - y \in K$, we have

$$\begin{aligned} p(x) &= \lim_{r \to 0^+} \sup_{r' \leq r} \frac{\mathcal{P}(B(x,r'))}{|B(x,r')|} \\ &\leq \lim_{r \to 0^+} \sup_{r' \leq r} \frac{e^{\varepsilon}\mathcal{P}(B(y,r'))}{|B(x,r')|} \\ &= e^{\varepsilon} \limsup_{r \to 0^+} \frac{\mathcal{P}(B(y,r'))}{|B(y,r')|} \\ &= e^{\varepsilon} p(y) \end{aligned}$$

Since K has positive measure, for every x there must exist a $y \in x - K$ with $p(y) < \infty$. Therefore p is finite everywhere. □

Next, we show that the density $p(x)$ can be modified to depend only on the sensitivity norm $\|x\|_K$ without affecting its privacy or utility.

Lemma 5 ("Polar" representation)**.** *There exists a parameterization $\theta \in \Theta$ of the unit sphere $\{u : \|u\|_K = 1\}$ under the mapping $u(\theta)$ and a function $g(\theta)$ such* *that*

$$\int f(z)\,d\lambda(z) = \int_\Theta \int_0^\infty f(ru(\theta))\, r^{d-1} g(\theta)\, dr\, d\theta$$

for any integrable function f.

Proof. See, for example, Blumenson [1960]. □

Now, define $\bar{p}(x)$ from p by averaging over vectors with the same norm:

$$\bar{p}(x) = \frac{1}{C} \int_\Theta p(\|x\|_K u(\theta))\, g(\theta)\, d\theta \ ,$$

where $C = \int_\Theta g(\theta)\, d\theta$.

Lemma 6. *$\bar{p}$ is pointwise ε-differentially private.*

Proof. For any $z \in \mathbb{R}^d$ and any $x \in K$,

$$\begin{aligned} \bar{p}(z) &= \frac{1}{C} \int_\Theta p(\|z\|_K u(\theta))\, g(\theta)\, d\theta \\ &\le \frac{1}{C} \int_\Theta e^\varepsilon p(\|z+x\|_K u(\theta))\, g(\theta)\, d\theta \\ &= e^\varepsilon \bar{p}(z+x) \ , \end{aligned}$$

since $|\|z\|_K - \|z+x\|_K| \le \|x\|_K \le 1$, and thus $\|z\|_K u(\theta) - \|z+x\|_K u(\theta) \in K$. □

Next we show that the expected loss, measured by the K-norm, is the same for $\bar{p}$ as for p.

Lemma 7. $\int \|z\|_K\, p(z)\, d\lambda(z) = \int \|z\|_K\, \bar{p}(z)\, d\lambda(z)$.

Proof.

$$\begin{aligned} &\int \|z\|_K\, p(z)\, d\lambda(z) \\ &= \int_\Theta \int_0^\infty \|ru(\theta)\|_K\, p(ru(\theta))\, r^{d-1} g(\theta)\, dr\, d\theta \\ &= \int_0^\infty r^d \int_\Theta p(ru(\theta)) g(\theta)\, d\theta\, dr \\ &= C \int_0^\infty r^d \bar{p}(ru)\, dr \\ &= \int_\Theta \int_0^\infty \|ru(\theta)\|_K\, \bar{p}(ru(\theta)) r^{d-1} g(\theta)\, dr\, d\theta \\ &= \int \|z\|_K\, \bar{p}(z)\, d\lambda(z) \ , \end{aligned}$$

where a bare u denotes any unit vector under $\|\cdot\|_K$. □

We have shown that any ε-DP additive noise mechanism can be replaced by a mechanism whose density $\bar{p}(x)$ depends only on $\|x\|_K$ without sacrificing privacy or utility. Thus, there must exist an optimal additive noise mechanism whose density depends only on the K-norm. (This establishes Theorem 1.) We can therefore restrict our search to such mechanisms; if a mechanism is optimal within this class, it is also optimal overall.

6.2 Optimality Of Generalized Staircase Mechanisms

We now sketch the argument for Theorem 2. Let $\mathcal{P}$ be an ε-differentially private noise distribution on $\mathbb{R}^d$ with density $p(x) = f(\|x\|_K)$. For $R \subseteq [0,\infty)$, let $\mathcal{P}(R)$ denote $\mathcal{P}(\{x : \|x\|_K \in R\})$ and let $|R|$ denote the Lebesgue measure of $\{x : \|x\|_K \in R\}$.

Definition 8. *The* upper radial density *of $\mathcal{P}$ is*

$$\hat{f}(r) := \limsup_{\delta \to 0^+} \frac{\mathcal{P}([r-\delta, r+\delta])}{|[r-\delta, r+\delta]|}$$

for $r > 0$, and $\hat{f}(0) = \limsup_{r\to 0^+} \hat{f}(r)$. This density is always defined and bounded, following the proof of Lemma 4, and fully determines $\mathcal{P}$ through the density $\hat{p}(x) = \hat{f}(\|x\|_K)$.

The upper radial density is a convenient representation of $\mathcal{P}$ because it has nice regularity properties. In particular, we have the following lemma.

Lemma 8. *Let $\hat{f}$ be the upper radial density of an ε-differentially private noise distribution $\mathcal{P}$. Then for any $r \in (0,\infty)$, $\delta_0 > 0$, and $\beta > 0$, there exists a $\delta' < \delta_0$ such that*

$$\hat{f}(r) - \beta < \mathcal{P}([r-\delta', r+\delta'])/|[r-\delta', r+\delta']| < \hat{f}(r) + \beta \ .$$

Proof. Expanding the definition of the upper radial density, we have

$$\hat{f}(r) = \lim_{\delta\to 0^+} \sup_{\delta' < \delta} \frac{\mathcal{P}([r-\delta', r+\delta'])}{|[r-\delta', r+\delta']|} \ .$$

From the definition of the limit, there exists $\delta < \delta_0$ such that the supremum is less than $\hat{f}(r) + \beta$, thus for all $\delta' < \delta$ we have $\mathcal{P}([r-\delta', r+\delta'])/|[r-\delta', r+\delta']| < \hat{f}(r) + \beta$. At the same time, the supremum is at least $\hat{f}(r)$, and so by the definition of supremum there exists some $\delta' < \delta$ such that $\mathcal{P}([r-\delta', r+\delta'])/|[r-\delta', r+\delta']| > \hat{f}(r) - \beta$. □

Lemma 8 allows us to translate pointwise properties of the density $\hat{f}$ to properties of positive probability masses under $\mathcal{P}$. This technical result is critical to proving the following three lemmas, which mirror the main steps in the argument of Geng et al. [2015]. (The proofs are provided in Appendix A.)

Lemma 9. *If $\mathcal{P}$ is a (K,ε)-optimal noise distribution, then its upper radial density $\hat{f}$ is monotone non-increasing.*

Lemma 10. *If $\mathcal{P}$ is a (K,ε)-optimal noise distribution, and $\hat{f}$ is its upper radial density, then for all $r \in [0,\infty)$, $\hat{f}(r) = e^\varepsilon \hat{f}(r+1)$.*

Lemma 11. *If $\mathcal{P}$ is a (K, ε)-optimal noise distribution, and $\hat{f}$ is its upper radial density, then for some $\gamma \in [0, 1]$ and some constant c we have*

$$\hat{f}(r) = \begin{cases} c & 0 < r \leq \gamma \\ e^{-\varepsilon} c & \gamma < r \leq 1 \end{cases}$$

By Lemma 17, a (K, ε)-optimal noise distribution exists. Combining Lemmas 10 and 11, we conclude that there exists an optimal generalized staircase distribution, establishing Theorem 2.

7 Future Directions

We showed that the generalized staircase mechanism is optimal (a) for convex sensitivity spaces, (b) under the sensitivity norm, and (c) among the class of additive noise mechanisms. Any of these qualifications could potentially be relaxed. Is there a general way to derive optimal mechanisms for sensitivity spaces that are not convex, as Kulesza et al. [2023] did in their specific setting? Can we derive optimal mechanisms for error metrics that do not match the sensitivity space? Is the staircase mechanism (or some other mechanism) optimal among all differentially private mechanisms, even those that are not additive?

Additionally, while we focused on pure differential privacy here, in practice relaxations such as approximate differential privacy or zero-concentrated differential privacy are commonly used. Extending the work of Geng et al. [2020] to derive optimal mechanisms under a relaxed privacy model in the general setting of arbitrary high-dimensional sensitivity spaces might yield significant practical benefits.

References

Kareem Amin, Jennifer Gillenwater, Matthew Joseph, Alex Kulesza, and Sergei Vassilvitskii. Plume: differential privacy at scale. *arXiv preprint arXiv:2201.11603*, 2022.

Hilal Asi and John C Duchi. Near instance-optimality in differential privacy. *arXiv preprint arXiv:2005.10630*, 2020.

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

LE Blumenson. A derivation of n-dimensional spherical coordinates. *The American Mathematical Monthly*, 67(1):63–66, 1960.

Mark Bun and Thomas Steinke. Average-case averages: Private algorithms for smooth sensitivity and mean estimation. In *Advances in Neural Information Processing Systems*, pages 181–191, 2019.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*, pages 265–284. Springer, 2006.

Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4): 211–407, 2014.

Quan Geng and Pramod Viswanath. The optimal mechanism in differential privacy. In *2014 IEEE international symposium on information theory*, pages 2371–2375. IEEE, 2014.

Quan Geng, Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The staircase mechanism in differential privacy. *IEEE Journal of Selected Topics in Signal Processing*, 9(7):1176–1184, 2015.

Quan Geng, Wei Ding, Ruiqi Guo, and Sanjiv Kumar. Tight analysis of privacy and utility tradeoff in approximate differential privacy. In *International Conference on Artificial Intelligence and Statistics*, pages 89–99. PMLR, 2020.

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

Matthew Joseph and Alexander Yu. Some constructions of private, efficient, and optimal k-norm and elliptic gaussian noise. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 2723–2766. PMLR, 2024.

Alex Kulesza, Ananda Theertha Suresh, and Yuyan Wang. Mean estimation in the add-remove model of differential privacy. *arXiv preprint arXiv:2312.06658*, 2023.

Ryan Rogers, Subbu Subramaniam, Sean Peng, David Durfee, Seunghyun Lee, Santosh Kumar Kancha, Shraddha Sahay, and Parvez Ahammad. Linkedin's audience engagements api: A privacy preserving data analytics system at scale. *arXiv preprint arXiv:2002.05839*, 2020.

Salil Vadhan. The complexity of differential privacy. *Tutorials on the Foundations of Cryptography: Dedicated to Oded Goldreich*, pages 347–450, 2017.

Jia Xu, Zhenjie Zhang, Xiaokui Xiao, Yin Yang, Ge Yu, and Marianne Winslett. Differentially private histogram publication. *The VLDB journal*, 22: 797–822, 2013.

Checklist

1. For all models and algorithms presented, check if you include:

   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]

   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]

   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [No]

2. For any theoretical claim, check if you include:

   (a) Statements of the full set of assumptions of all theoretical results. [Yes]

   (b) Complete proofs of all theoretical results. [Yes] All proofs are provided in the appendix.

   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:

   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes] Instructions to reproduce experiments are provided.

   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable]

   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]

   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes] All experiments are run on CPUs.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:

   (a) Citations of the creator If your work uses existing assets. [Not Applicable]

   (b) The license information of the assets, if applicable. [Not Applicable]

   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]

   (d) Information about consent from data providers/curators. [Not Applicable]

   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

   (a) The full text of instructions given to participants and screenshots. [Not Applicable]

   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

A ANALYSIS

A.1 Proofs Of Lemma 1 and Corollary 1

Recall that

$$p(x) \propto e^{-\lfloor \|x\|_K - \gamma \rfloor \varepsilon}.$$

Let C be the normalization constant and $f_\gamma(r) = \frac{1}{C} e^{-\lfloor r-\gamma \rfloor \varepsilon}$. Then, the expected loss can be written as

$$\mathbb{E}_{X \in \mathcal{P}} \left[\|X\|_K \right] = \int_{x \in \mathbb{R}^d} p(x) \|x\|_K dx = \int_{x \in \mathbb{R}^d} f_\gamma(\|x\|_K) \|x\|_K dx$$

By Lemma 5, this can be rewritten as

$$\mathbb{E}_{X \in \mathcal{P}} \left[\|X\|_K \right] = \int_\Theta \int_0^\infty f_\gamma(r) r r^{d-1} g(\theta) dr d\theta = \int_0^\infty f_\gamma(r) r^d dr \int_\Theta g(\theta) d\theta.$$

Similarly, since $\mathcal{P}$ is a distribution,

$$\mathbb{E}_{X \in \mathcal{P}} \left[1 \right] = \int_\Theta \int_0^\infty f_\gamma(r) r^{d-1} g(\theta) dr d\theta = \int_0^\infty f_\gamma(r) r^{d-1} dr \int_\Theta g(\theta) d\theta.$$

Taking the ratio of the above two equations and using the fact that $\mathbb{E}_{X \in \mathcal{P}} [1] = 1$ yields,

$$\mathbb{E}_{X \in \mathcal{P}} \left[\|X\|_K \right] = \frac{\int_0^\infty f_\gamma(r) r^d dr}{\int_0^\infty f_\gamma(r) r^{d-1} dr},$$

which is independent of K. Hence both optimal loss and the optimal value of γ are independent of K.

A.2 Proofs Of Theorems 3 and 4

Here we discuss in detail how the general staircase mechanism's utility changes compared to the K-norm in the two regimes mentioned in 3.2, and prove Theorems 3 and 4.

For the convenience of writing proofs, we introduce some new notations. Let $C_{d,\varepsilon}(\gamma) = \sum_{i=0}^{+\infty} (i+\gamma)^d e^{-\varepsilon i}$ be the normalizing constant in the discrete gamma distribution $D\Gamma_{\gamma,d,\varepsilon}$. Denote the staircase distribution with any γ value by $\mathcal{P}_\gamma$. Let $h_{d,\varepsilon}(\gamma) = \mathbb{E}_{X \in \mathcal{P}_\gamma} [\|X\|_K]$ be the expected error at any γ. Notice that it is written as a function of γ. In Lemma 1, we have shown $h_{d,\varepsilon}(\gamma)$ is independent of sensitivity space K, so for the purposes of this analysis, we can treat K to be an $\|\cdot\|_2$ unit ball. The next lemma bounds the expected ℓ_2-norm of a random sample from unit ball.

Lemma 12 (Folklore). *If X is a uniform random sample from the ℓ_2 unit ball in $\mathbb{R}^d$, then the expected ℓ_2-norm of X is $\frac{d}{d+1}$.*

We next relate $h_{d,\varepsilon}(\gamma)$ to the normalizing constants in the discrete gamma distribution.

Lemma 13. *For any d, $\varepsilon > 0$, and $\gamma \in [0,1]$,*

$$h_{d,\varepsilon}(\gamma) = \frac{d}{d+1} \cdot \frac{C_{d+1,\varepsilon}(\gamma)}{C_{d,\varepsilon}(\gamma)}$$

Proof. Consider Algorithm 2 which samples from the staircase distribution. Conditioning on that the sampled scale $y = i+\gamma$ for any non-negative integral i, the expected norm value of the algorithm's output xy is $\frac{d}{d+1}(i+\gamma)$. Hence

$$h_{d,\varepsilon}(\gamma) = \frac{d}{d+1} \sum_{i=0}^\infty (i+\gamma) \cdot \frac{(i+\gamma)^d e^{-\varepsilon i}}{C_{d,\varepsilon}(\gamma)} = \frac{d}{d+1} \cdot \frac{C_{d+1,\varepsilon}(\gamma)}{C_{d,\varepsilon}(\gamma)}$$

□

The next lemma characterizes the optimal γ, denoted as $\gamma^*(\varepsilon, d)$ for staircase mechanism.

Lemma 14. *For $d \geq 1$, $\gamma^*(\varepsilon, d) \in [0,1]$ satisfies the following equality:*

$$(d+1) \cdot C_{d,\varepsilon}^2(\gamma^*(\varepsilon,d)) = d \cdot C_{d+1,\varepsilon}(\gamma^*(\varepsilon,d)) \cdot C_{d-1,\varepsilon}(\gamma^*(\varepsilon,d)) \tag{1}$$

Proof. Taking the derivative of h with respect to γ yields:

$$\frac{\partial h}{\partial \gamma} = \frac{d}{d+1} \cdot \frac{(d+1)C_{d,\varepsilon}^2(\gamma) - dC_{d+1,\varepsilon}(\gamma)C_{d-1,\varepsilon}(\gamma)}{C_{d,\varepsilon}^2(\gamma)}. \tag{2}$$

Let $\gamma^* \triangleq \gamma^*(\varepsilon, d)$ for notational simplicity. We argue that gradient at γ^* has to be zero and hence the result follows by Equation 2.

Suppose otherwise. Since the gradient at γ^* is not zero, the only possible values of γ^* are zero and one. However, note that $\forall d \geq 1, \varepsilon$, $C_{d,\varepsilon}(0) = \sum_{i=1}^{\infty} i^d e^{-\varepsilon i}$, which is same as $e^{-\varepsilon} C_{d,\varepsilon}(1)$; and we have that $\forall d, \varepsilon$, $h_{d,\varepsilon}(0) = h_{d,\varepsilon}(1)$, and $h'(d,\varepsilon)(0) = h'(d,\varepsilon)(1)$. Suppose $h'(d,\varepsilon)(0) < 0$, then, for a sufficiently small value of γ, $h(\gamma) < h(0)$. Hence a contradiction. Similarly, if $h'(d,\varepsilon)(0) > 0$, then $h'(d,\varepsilon)(1) > 0$ and for a sufficiently small value of γ, $h(1-\gamma) < h(1)$. Hence a contradiction. □

We state two simple results on the properties of continuous K-norm mechanisms.

Lemma 15. *The error of the K-norm mechanism is independent of the underlying sensitivity space K and depends only on the privacy parameter ε and the dimension d.*

The proof of the above result is similar to that of Corollary 1 and omitted. The above lemma together in conjunction with the fact that when K-norm is ℓ_1 norm, K-norm mechanism is same as Laplace mechanism, implies the following corollary.

Corollary 2. *Let $\varepsilon > 0$. For any K-norm in d-dimensions, the error of the K-norm mechanism is given by*

$$\mathbb{E}[\|X\|_K] = \frac{d}{\varepsilon}.$$

Using the above lemma, we prove the following bound on ratio of optimal losses for consecutive values of d.

Lemma 16. *Let $h^*(d,\varepsilon) = \min_{\gamma \in [0,1]} h_{d,\varepsilon}(\gamma)$. For any $d \geq 1$ we have $\frac{h^*(d,\varepsilon)}{h^*(d-1,\varepsilon)} \geq \frac{d}{d-1}$.*

Proof. Plug (1) into the expression $h_{d,\varepsilon}(\gamma^*(\varepsilon,d))$:

$$\begin{aligned} h^*(d,\varepsilon) &= \frac{d}{d+1} \cdot \frac{d+1}{d} \cdot \frac{C_{d,\varepsilon}(\gamma^*(\varepsilon,d))}{C_{d-1,\varepsilon}(\gamma^*(\varepsilon))} \\ &= \frac{d}{d-1} h_{d-1,\varepsilon}(\gamma^*(d,\varepsilon)) \\ &\geq \frac{d}{d-1} h^*(d-1,\varepsilon) \end{aligned}$$

□

We now have all the tools to prove Theorems 3 and 4.

Proof of Theorem 3. Let $h_c(d,\varepsilon)$ denote the error of the continuous K-norm mechanism in d dimensions. Combining Lemma 16 and Corollary 2 yields

$$\frac{h^*(d,\varepsilon)}{h^*(d-1,\varepsilon)} \geq \frac{d}{d-1} = \frac{h_c(d,\varepsilon)}{h_c(d-1,\varepsilon)}.$$

Rearranging terms, we get

$$r(\varepsilon,d) = \frac{h^*(d,\varepsilon)}{h_c(d,\varepsilon)} \geq \frac{h^*(d-1,\varepsilon)}{h_c(d-1,\varepsilon)} = r(\varepsilon,d-1),$$

and hence the theorem. □

Proof of Theorem 4. We first prove the result for $\varepsilon \to 0$. By Lemma 16,

$$h^*(d, \varepsilon) \geq dh^*(1, \varepsilon).$$

By Geng and Viswanath [2014, Corollary 5],

$$h^*(1, \varepsilon) \geq \frac{1}{\varepsilon}\left(1 - o(1)\right).$$

Combining the above two equations, together with Corollary 2 yields

$$h^*(d, \varepsilon) \geq dh^*(1, \varepsilon) \geq \frac{d}{\varepsilon}\left(1 - o(1)\right) = h_c(d, \varepsilon)\left(1 - o(1)\right),$$

where $h_c(d, \varepsilon)$ is the error of the d-dimensional continuous K-norm mechanism with privacy parameter ε. Hence, as $\varepsilon \to 0$, $r(\varepsilon, d) \to 1$.

We now focus on the regime where ε is large. In this regime, observe that

$$\begin{aligned}
\min_{\gamma \in [0,1]} h_{d,\varepsilon}(\gamma) &\leq h_{d,\varepsilon}(e^{-\varepsilon/(d+1)}) \\
&= \frac{d}{d+1} \frac{C_{d+1,\varepsilon}(e^{-\varepsilon/(d+1)})}{C_{d,\varepsilon}(e^{-\varepsilon/(d+1)})} \\
&= \frac{d}{d+1} \frac{O(e^{-\varepsilon(d+1)/(d+1)})}{O(e^{-\varepsilon(d)/(d+1)})} \\
&= O(e^{-\varepsilon/(d+1)}).
\end{aligned}$$

The proof follows by combining the above bound with Corollary 2. □

A.3 Analysis Of Implementation

In this subsection we focus on analyzing Algorithm 1 and 2. We first prove Lemma 2 which shows Algorithm 1 samples from the correct discrete gamma distribution $D\Gamma_{\gamma,\varepsilon,d}$.

Proof of Lemma 2. The last step of the algorithm is a rejection-sampling procedure. Upon each new sampling iteration, let $Z = \lceil s - \gamma \rceil + \lfloor Y - t \rfloor$ be the random variable denoting the sampling outcome. the probability of returning a particular y is:

$$\begin{aligned}
\mathbb{P}[Z = i] &= \sum_{z=0}^{i} \mathbb{P}[\sum_{i=0}^{d} \lceil s - \gamma \rceil = z] \cdot \mathbb{P}[\lfloor Y - t \rfloor = i - z] \\
&= \sum_{z=0}^{i} \int_{\min\{z-1,0\}+\gamma \leq x \leq z+\gamma} p(s = x) \cdot \mathbb{P}[\lfloor Y - t \rfloor = i - z].
\end{aligned}$$

For any $s \in [\min\{z - 1, 0\} + \gamma, z + \gamma$, we compute the probability of the matching Y value: $\mathbb{P}[\lfloor Y - t \rfloor = i - z]$, where t, as defined in Algorithm 1, is $\lceil s - \gamma \rceil + \gamma - s$ which always guarantees $t \in [0, 1]$. Hence this probability

$$\mathbb{P}[\lfloor Y - t \rfloor = i - z] = \int_{y=(i-z)+t}^{(i-z+1)+t} e^{-\varepsilon y} = e^{-\varepsilon(i-z+t)} \cdot (1 - e^{-\varepsilon}) \propto e^{-\varepsilon(i-z+t)},$$

where the $\propto$ sign leaves out constants. Notice that given any x such that $\lceil x - \gamma \rceil = z$, when $s = x$, by definition of t,

$$e^{-\varepsilon s} \cdot e^{-\varepsilon(i-z+t)} \equiv e^{-\varepsilon(i+\gamma)}.$$

Plugging this into the previous equation we have

$$\begin{aligned}
\mathbb{P}[Z=i] &= \sum_{z=0}^{i} \int_{\min\{z-1,0\}+\gamma \le x \le z+\gamma} p(s=x)\cdot \mathbb{P}[\lfloor Y-t \rfloor = i-z] \\
&\propto \sum_{z=0}^{i} \int_{\min\{z-1,0\}+\gamma \le x \le z+\gamma} e^{-\varepsilon x}\cdot \mathbb{P}[\lfloor Y-t \rfloor = i-z] \\
&\propto \sum_{z=0}^{i} \int_{\min\{z-1,0\}+\gamma \le x \le z+\gamma} 1\cdot e^{-\varepsilon i} \\
&= e^{\varepsilon i} \int_{\sum_{j=1}^{d} X_j = 0}^{i+\gamma} 1 \\
&\propto (i+\gamma)^d e^{-\varepsilon i},
\end{aligned}$$

where in the last line we used the fact that:

$$\int_{X_1=0}^{x} \int_{X_2=0}^{x-X_1} \int_{X_3=0}^{x-(X_1+X_2)} \ldots dX_d dX_{d-1} \cdots dX_1 = \frac{x^d}{d!}$$

holds for any $x \ge 0$. □

We now prove the correctness of Algorithm 2 using the fact that Algorithm 1 returns a sample from the discrete gamma distribution.

Proof of Theorem 5. Pick any point $v \in \mathbb{R}^d$ such that $\lfloor \|v\|_K \rfloor \in [(i-1)+\gamma, i+\gamma)$ $(i=1,2,\ldots)$. For Algorithm 2 to sample v, the scalar y that it samples from $D\Gamma_{\gamma,\varepsilon,d}$ must be at least $i+\gamma$, otherwise we must have $\|xy\| \le y\|x\|_K \le (i-1)+\gamma$. When $y=j+\gamma$, the probability of sampling any x within K is $\frac{1}{V(K)}$, where $V(K)$ is the volume of K. Hence

$$\begin{aligned}
p(xy=v) &= \sum_{j=i}^{+\infty} \mathbb{P}[y=j]\cdot p\left(x=\frac{v}{y}\right) \\
&\propto \sum_{j=i}^{+\infty} (j+\gamma)^d e^{-\varepsilon j}\cdot \frac{1}{(j+\gamma)^d} \\
&= \sum_{j=i}^{+\infty} e^{\varepsilon j} \propto e^{-\varepsilon i}.
\end{aligned}$$

Hence it samples from the correct staircase distribution.

To prove Theorem 5 it remains only to discuss the run-time of the sampling algorithm. To output one sample, it only makes one call to the given uniform-sampling subroutine $\mathcal{U}$ on K, hence we only need to show it can also sample y in $O(d)$ time for fixed ε and γ. Suppose we have already sampled $X_1,\ldots,X_d$, and obtained s and t that is how far s is from the next $i+\gamma$ for non-negative integral i. t is guaranteed to be in $[0,1)$. Given any t the probability of rejecting the sample is at most $\mathbb{P}[Y \le t] \le 1-e^{-\varepsilon}$. Hence the algorithm is expected to return a valid sample in $O(1)$ calls to the rejection sampling process, each call taking $O(d)$ time to sample all the random variables. □

A.4 Proofs Of Lemmas 9, 10, and 11

We first prove that the optimal noise distribution exists.

Lemma 17. *There exists an optimal (K,ε)-differentially private noise distribution $\mathcal{P}^*$.*

Proof. Let $\mathbb{P}$ denote the set of all probability distributions. The (K,ε)-differential privacy constraints can be expressed as a collection of linear inequalities in terms of $\mathcal{P}$. Hence, each differential privacy constraint defines a closed set in the space of probability distributions. By De Morgan's laws, the intersection of closed sets, even an infinite number of them, remains a closed set. Therefore, the set of all (K, ε)-differentially private distributions, denoted by $\mathbb{P}_{DP}$, is a closed subset of $\mathbb{P}$.

The objective function $\mathbb{E}_{Z\sim\mathcal{P}}\left[\|Z\|_K\right]$ is continuous and linear with respect to $\mathcal{P}$ and since the norm is always non-negative, $\inf_{\mathcal{P}\in\mathbb{P}} \mathbb{E}_{Z\sim\mathcal{P}}\left[\|Z\|_K\right]$ is bounded below by zero. Hence, the minimization of $\mathbb{E}_{Z\sim\mathcal{P}}\left[\|Z\|_K\right]$ over $\mathbb{P}_{DP}$ is a minimization of a continuous function over a closed set with bounded infimum. Hence the infimum is attained, and thus, an optimal noise distribution $\mathcal{P}^*$ exists. □

Before proving these three lemmas, suppose we are given a radial density function $\hat{f}$. Let $V(r)$ denote the volume of the set $\{x : \|x\|_K \le r\}$. $\hat{f}$ is a valid probability function if: $\int_{r=0}^{\infty} \hat{f}(r)dV(r) = 1$. Also, we assume that the Riemann integral $\mathbb{E}_{x\sim\hat{f}_i}[\|x\|_K] = \int_{r=0}^{\infty} \hat{f}(r)rdV(r)$ exists, which is the expected loss.

We first introduce a series of discretized probability functions whose utility converges to the given $\hat{f}$. Let i be an integer used to discretize $\hat{f}$; for each i, define a discretized probability function $\hat{f}_i$ to be on the support of the set $\{\frac{k}{i}\}_{k=0}^{\infty}$:

$$\hat{f}_i\left(\frac{k}{i}\right) = c'(i)\hat{f}\left(\frac{k}{i}\right), \forall k \in \mathbb{Z}^+$$

Notice that $\mathbb{P}[\frac{k}{i}] = \hat{f}_i\left(\frac{k}{i}\right)\cdot(V\left(\frac{k+1}{i}\right) - V\left(\frac{k}{i}\right))$, where $V\left(\frac{k+1}{i}\right) - V\left(\frac{k}{i}\right)$ is the volume of the set: $\{v : \|v\|_K \in (\frac{k}{i}, \frac{k+1}{i})\}$, and the constant $c'(i)$ is the normalizing term used to guarantee that $\hat{f}_i$ is a valid probability that satisfies $\sum_k \hat{f}_i\left(\frac{k}{i}\right)(V\left(\frac{k+1}{i}\right) - V\left(\frac{k}{i}\right)) = 1$:

$$c'(i) = \frac{1}{\sum_{k=0}^{\infty} \hat{f}\left(\frac{k}{i}\right)\cdot(V\left(\frac{k+1}{i}\right) - V\left(\frac{k}{i}\right))}$$

Lemma 18. *The series of functions $\{\hat{f}_i\}_{i=1}^{\infty}$ are valid probability functions that guarantees the total probability mass sums up to 1, and their expected utility converges to that of $\hat{f}$.*

Proof. Notice that the value $c'(i)$ always exists, since ε-DP $\hat{f}$ satisfies:

$$e^{-\varepsilon}\sum_{k=0}^{\infty}\hat{f}\left(\frac{k}{i}\right)\left(V\left(\frac{k+1}{i}\right) - V\left(\frac{k}{i}\right)\right)) \le \int_{r=0}^{\infty}\hat{f}(r)dV(r) \le e^{\varepsilon}\sum_{k=0}^{\infty}\hat{f}\left(\frac{k}{i}\right)\left(V\left(\frac{k+1}{i}\right) - V\left(\frac{k}{i}\right)\right)$$

and we also have $\lim_{i\to\infty} c'(i) = 1$, since

$$\lim_{i\to\infty}\sum_k \hat{f}\left(\frac{k}{i}\right)\cdot\left(V\left(\frac{k+1}{i}\right) - V\left(\frac{k}{i}\right)\right) = \int_{r=0}^{\infty}\hat{f}(r)dV(r) = 1$$

This is true because the left-hand side is the Riemann Sum of the integral on the right hand side.

Each $\hat{f}_i(\cdot)$ is discrete and monotone non-increasing, and its utility $\mathbb{E}_{x\sim\hat{f}_i}[\|x\|_K]$ always exists. Further, by definition of the $\hat{f}_i$'s, the limit is:

$$\begin{aligned}
\lim_{i\to\infty}\mathbb{E}_{x\sim\hat{f}_i}[\|x\|_K] &= \lim_{i\to\infty} c'_i\cdot\sum_{k=0}^{\infty}\hat{f}\left(\frac{k}{i}\right)\cdot\frac{k}{i}\cdot\left(V\left(\frac{k+1}{i}\right) - V\left(\frac{k}{i}\right)\right)\\
&= \lim_{i\to\infty} c'_i\cdot\lim_{i\to\infty}\sum_{k=0}^{\infty}\hat{f}\left(\frac{k}{i}\right)\cdot\frac{k}{i}\cdot\left(V\left(\frac{k+1}{i}\right) - V\left(\frac{k}{i}\right)\right)\\
&= 1\cdot\int_{r=0}^{\infty}\hat{f}(r)rdV(r)\\
&= \mathbb{E}_{x\sim\hat{f}}[\|x\|_K]
\end{aligned}$$

Where we used the fact that $\sum_k \hat{f}\left(\frac{k}{i}\right)\cdot\frac{k}{i}\cdot(V\left(\frac{k+1}{i}\right) - V\left(\frac{k}{i}\right))$ is the Riemann sum of $\int_{r=0}^{\infty}\hat{f}(r)rdV(r)$. □

Fix any integer i and let $\mathcal{F}_{\varepsilon,i}$ be the class of all such discrete probability functions f such that it only has support on discrete set $\{\frac{k}{i}\}_{k\in\mathbb{Z}^+}$, and $\sum_k f\left(\frac{k}{i}\right)\cdot\left(V\left(\frac{k+1}{i}\right)-V\left(\frac{k}{i}\right)\right)=1$, and also satisfies the privacy constraint: $\frac{f(i)}{f(j)}\in[e^{-\varepsilon},e^{\varepsilon}]$ if $|i-j|\leq 1$. Notice that $\hat{f}_i\in\mathcal{F}_{\varepsilon,i}$. We first show that Lemma 9, 10, 11 hold within the scope of $\mathcal{F}_{\varepsilon,i}$. That is, the following claim always holds for any i.

We first state that Lemma 9, 10 and 11 must hold for any optimal distribution that has support on $\{\frac{k}{i}\}_{k\in\mathbb{Z}^+}$, for any given i.

Lemma 19. *For any fixed $i\in\mathbb{Z}^+$, there exists a function $\hat{f}_i^*(\cdot)$ that defines a probability distribution whose support is the set $\{\frac{k}{i}\}_{k\in\mathbb{Z}^+}$, satisfying*

$$\sum_{k=0}^{\infty}\hat{f}_i^*(\cdot)\cdot\left(V\left(\frac{k+1}{i}\right)-V\left(\frac{k}{i}\right)\right)=1$$

that minimizes the expected K-norm value among all such discrete probabilities, and satisfy the following properties:

1. *It is monotone non-increasing.*
2. *It declines cyclically: $\hat{f}_i^*\left(\frac{k}{i}\right)=e^{\varepsilon}\hat{f}_i^*(1+\frac{k}{i})$ always holds.*
3. *It is stair-shaped like stated in Lemma 11.*

We prove the three lemmas one by one.

Lemma 9. *If $\mathcal{P}$ is a (K,ε)-optimal noise distribution, then its upper radial density $\hat{f}$ is monotone non-increasing.*

Proof. Pick an optimal $\mathcal{P}\in\mathcal{F}_{\varepsilon,i}$, the probability function is $f(\cdot)$ defined on $\{\frac{k}{i}\}$. Suppose for contradiction there is $p<q\in\mathbb{Z}^+$ such that $f(\frac{p}{i})<f(\frac{q}{i})$. Consider increasing $f(\frac{p}{i})$ and decrease by some very small mass probability and construct a new f':

$$f'\left(\frac{k}{i}\right)=\begin{cases}f(\frac{p}{i})+(V(\frac{q+1}{i})-V(\frac{q}{i}))\delta & k=p\\ f(\frac{q}{i})-(V(\frac{p+1}{i})-V(\frac{p}{i}))\delta & k=q\\ f\left(\frac{k}{i}\right) & k\neq p,q\end{cases}$$

Where δ is a very small but positive number. If new f' does not violate any ε-DP constraints, since it must have smaller expected loss than f, it contradicts the optimality assumption of f. Hence some constraint is violated either because $f'(\frac{p}{i})$ is bigger or $f'(\frac{q}{i})$ is smaller. There could be two cases:

Case 1: Suppose the ε-DP constraint is violated at $\frac{q}{i}$. There must exist some s, such that $|q-s|\leq i$, $f(\frac{q}{i})=e^{\varepsilon}f(\frac{p}{i})$. We show that $s>p$. Otherwise, $s<p<q$, so $p-s\leq q-s\leq m$; but we also have $f(\frac{s}{i})=e^{\varepsilon}f(\frac{q}{i})>e^{\varepsilon}f(\frac{p}{i})$, which violates the ε-DP constraint for s,p. Hence $s>p$, we have a new pair of numbers $p<s$, such that $f(\frac{p}{r})<f(\frac{s}{r})$, while the higher value of the two, $f(\frac{s}{r})$, is bigger than the previous $f(\frac{q}{r})$ by e^{ε}.

Case 2: The ε-DP constraint is violated at $\frac{p}{i}$. There must exist some t, such that $|p-t|\leq i$, $f(\frac{p}{i})=e^{\varepsilon}f(\frac{t}{i})$. Similarly, we can show that $t<q$. Hence we will find a new pair (t,q) such that $\frac{f(\frac{q}{i})}{f(\frac{t}{i})}>1$, with reduced value of the lower end $f(\frac{t}{i})<f(\frac{p}{i})$ and the higher value in the pair remains $f(\frac{q}{i})$.

Iteratively repeat this argument and we show that eventually we will be able to find a pair of integer values (x,y) such that we could construct a function with better utility by moving a small amount of probability mass from $f(\frac{y}{i})$ to $f(\frac{x}{i})$. Assume for contradiction that this process does not terminate and we can create an infinitely long series of such pairs $\{(x_t,y_t)\}_{t=0}^{\infty}$. Notice that in such a sequence, the function value at the higher end, $f(\frac{y}{i})$, must keep increasing from time to time. That is, for any pair (x_{t_1},y_{t_1}), $\exists(x_{t_2},y_{t_2})$, where $t_2>t_1$, and $f(\frac{y_{t_2}}{i})\geq e^{\varepsilon}f(\frac{y_{t_1}}{i})$. Otherwise it must be that $\forall t_2>t_1$, $y_{t_2}=y_{t_1}$; and the lower end, $f(\frac{x_{t_2}}{i})$, keeps shrinking by e^{ε} infinitely. This is impossible since if y is fixed at y_{t_1}, x cannot go beyond x_{t_1} and there are only finite number of values smaller than x_{t_1}.

Hence we must have an infinite sequences of $\{y\}$'s, the f values on these y's keeps increasing by e^{ε} every time. This contradicts the ε-DP constraint, which implies that the probabilities must be upper-bounded. Therefore, for f to be optimal, it must be monotone non-decreasing. □

Lemma 10. *If $\mathcal{P}$ is a (K, ε)-optimal noise distribution, and $\hat{f}$ is its upper radial density, then for all $r \in [0, \infty)$, $\hat{f}(r) = e^{\varepsilon} \hat{f}(r+1)$.*

Proof. The proof of this lemma is essentially derived in the same way as in Geng and Viswanath [2014], Appendix B, Step 5 in the proof of Theorem 3. □

Now we can focus on proving that there exists a stair-shaped $\hat{f}_i^*$ that is optimal.

Lemma 11. *If $\mathcal{P}$ is a (K, ε)-optimal noise distribution, and $\hat{f}$ is its upper radial density, then for some $\gamma \in [0, 1]$ and some constant c we have*

$$\hat{f}(r) = \begin{cases} c & 0 < r \leq \gamma \\ e^{-\varepsilon} c & \gamma < r \leq 1 \end{cases}$$

Proof. Here we only focus on proving that there exists a stair-shaped $\hat{f}_i^*$ that is optimal among the class of discretized probability functions $\mathcal{F}_{\varepsilon,i}$. We give a different interpretation of this claim. We call it a *short stair* for $\hat{f}_i^*$ if there exists some p, q, t, such that $p < q < t \leq p + i$, and also

$$\hat{f}_i^* \left(\frac{p}{i}\right) = \hat{f}_i^* \left(\frac{p+1}{i}\right) = \ldots = \hat{f}_i^* \left(\frac{q-1}{i}\right) > \hat{f}_i^* \left(\frac{q}{i}\right) = \hat{f}_i^* \left(\frac{q+1}{i}\right) = \ldots = \hat{f}_i^* \left(\frac{t-1}{i}\right) > \hat{f}_i^* \left(\frac{t}{i}\right)$$

(notice the strictly greater sign here). This claim is essentially equivalent to saying that there are no such short stairs for $\hat{f}_i^*$. For contradiction, say there is such a group of p, q, t which form a short stair. Consider two probabilities $\hat{f}_i^1$ and $\hat{f}_i^2$, both monotone and cyclically declining by e^{ε}, constructed as follows:

$$\hat{f}_i^1 \left(\frac{k}{i}\right) = \begin{cases} c_1 \hat{f}_i^* \left(\frac{k}{i}\right) & k \not\equiv q \mod i \\ c_1 \hat{f}_i^* (\frac{k-(q-p)}{i}) & k \equiv q \mod i \end{cases}$$

Where c_1 is a rescaling constant used to guarantee $\hat{f}_i^1$ is a valid probability function. Likewise,

$$\hat{f}_i^2 \left(\frac{k}{i}\right) = \begin{cases} c_2 \hat{f}_i^* \left(\frac{k}{i}\right) & k \not\equiv q \mod i \\ c_2 \hat{f}_i^* (\frac{k+(t-q)}{i}) & k \equiv q \mod i \end{cases}$$

Notice that, both $\hat{f}_i^1$ and $\hat{f}_i^2$ are monotone and acyclic, but we have removed the short stair p, q, t in both. In $\hat{f}_i^1$ we increase the probability density at $\frac{q}{i}$ to be equal to $\frac{p}{i}$; in $\hat{f}_i^2$ we decrease it to be equal to $\frac{t}{i}$. There exists $\lambda \in (0, 1)$, such that $\hat{f}_i^* = \lambda \hat{f}_i^1 + (1 - \lambda) \hat{f}_i^2$. Hence either $\hat{f}_i^1$ or $\hat{f}_i^2$ must have no greater loss than $\hat{f}_i^*$. We can keep doing this until we have removed all the short stairs. Hence we have proved that there exists some $\hat{f}_i^*$ that is both optimal and stair-shaped. □

Since we have proved Lemma 9,10 and 11 for the discrete function class $\mathcal{F}_{\varepsilon,i}$ we now generalize them to the continuous class, and prove that these lemmas hold for the (K, ε)-optimal continuous noise distribution.

Proof. For any given *optimal* $\hat{f}$, consider the constructed series of distributions $\{\hat{f}_i\}_{i=1,2,\ldots}$. For every i, in the class of all discrete distributions with the same support as $\hat{f}_i$, there exists an optimal one that is monotone, cyclically decreasing by e^{ε} every time $\|\cdot\|_K$ grows by 1, and is staircase-shaped with some $\gamma_i^* = \frac{m}{i}$ for some integer $m \in [i]$. Hence we have a sequence of optimal discrete mechanisms $\{\hat{f}_i^*\}_{i \in \mathbb{Z}^+}$, where for all i we have $\hat{f}_i^* \leq \hat{f}_i$. Every $\hat{f}_i^*$ is a discrete staircase with some γ_i^*, the sequence $\{\gamma_i^*\}$ is bounded, hence it must have a subsequence that converge to some value $\gamma \in [0, 1]$. Without loss of generality we just assume the subsequence is simply the whole sequence $\{\gamma_i^*\}$. Then, consider f_γ that is a continuous staircase distribution with stair length γ in $[0, 1]$. One can see that the sequences of utilities of $\{\gamma_i^*\}$ also converges to the utility of f_γ. Since every $\{\hat{f}_i^*\}$ has utility at least as good as $\{\hat{f}_i\}$ that is dicretized from $\hat{f}$, the utility value that they converge to must be at least as good as the utility of $\hat{f}$. Hence we have found a staircase distribution f_γ that is optimal. □

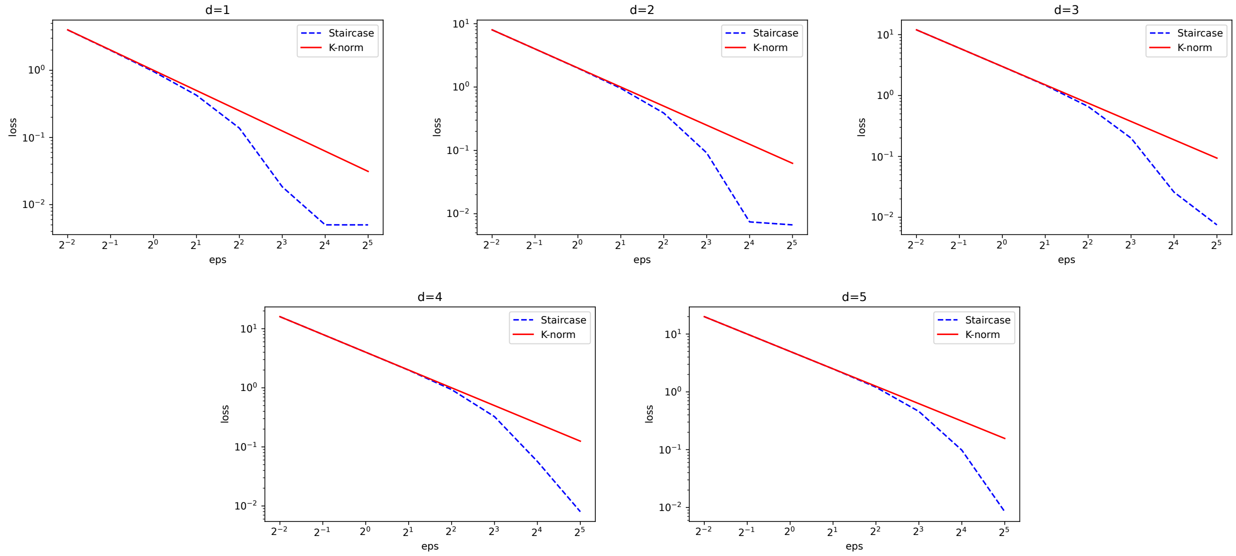

Figure 9: Expected error measured in $\| \cdot \|_K$ of staircase/K-norm mechanisms v.s. changing ε, fixed d

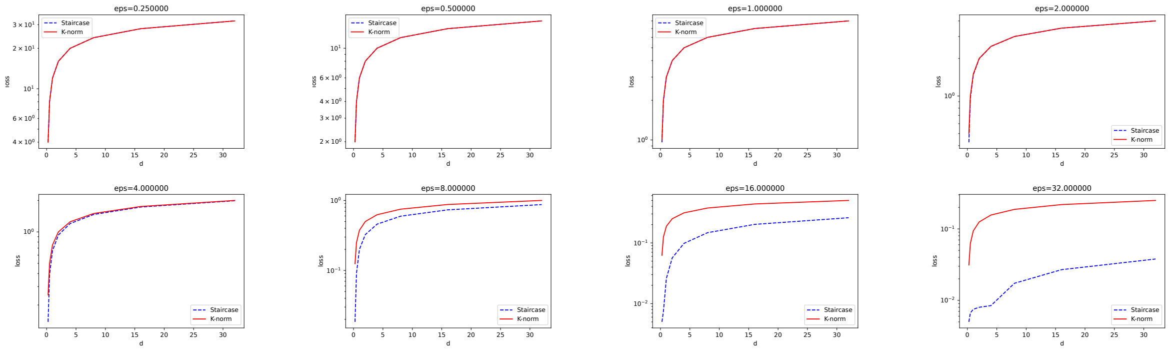

Figure 10: Expected error measured in $\| \cdot \|_K$ of staircase/K-norm mechanisms v.s. changing d, fixed ε

B ADDITIONAL EXPERIMENTS

In this section we include more experimental results. We show how ε (or d) affects the generalized staircase mechanism and the K-norm mechanism, and how the optimal γ value as discussed in Section 3 changes. Then we focus on showing that the generalized staircase mechanism outperforms the K-norm mechanism, as well as the Laplace mechanism, on different sensitivity spaces, for both matching and mismatching target norm error metrics.

B.1 Error In Different Parameter Settings

Here we further explore how the utilities of staircase and K-norm changes in different parameter settings. Among the three parameters (ε, d, γ) that can affect the utility of staircase, we assume γ is always chosen to minimize the expected error. In our implementation it is found via a grid search over the interval $[0, 1]$.

Figure 9 compares general staircase error with vanilla K-norm error while ε changes among $\{2^k\}_{k=-2}^{5}$, with d fixed to one of the values in $\{1, 2, 3, 4, 5\}$. The results further supports the conclusion that staircase outperforms K-norm in high-privacy regimes (bigger ε); whereas in low-privacy regimes (smaller ε) their performance tend to be the same. On the contrary, Figure 10 fixes ε at different values and change d. Indeed the error of staircase seems to be increasing faster than K-norm for fixed ε when d increases from 1 to 5.

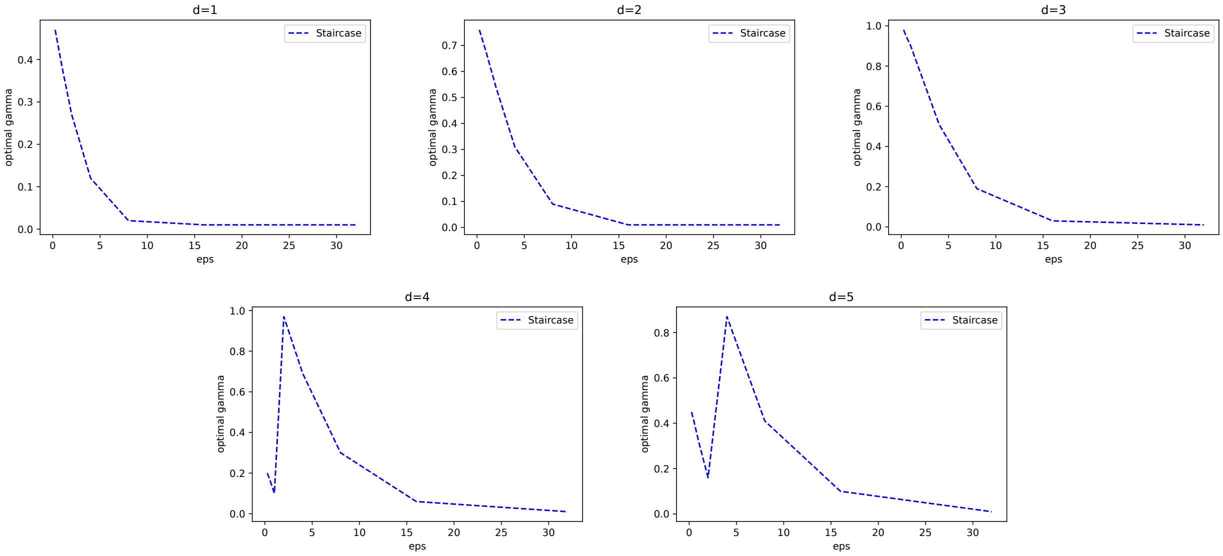

Figure 11: Optimal γ value, found by grid search over $[0, 1]$, v.s. different ε values, at fixed $d = 1, 2, 3, 4, 5$

B.2 Analysis of γ

Figure 11 shows how the value of the optimal γ changes, while fixing d and increasing ε. One can see that γ tends to converge to 0 in high-privacy regimes, but might not be monotonically decreasing depending on the d value.

Figure 12 shows how different choices of $\gamma \in [0, 1]$ affects the error of staircase mechanism in different scenarios of (ε, d). While the error can oscillate as γ changes from 0 to 1, there is often a wide range of γ values with which the staircase mechanism can outperform K-norm. Further, the error tends to be less curved around the optimal γ, showing the mechanism to be robust against choice of γ.

B.3 Performance For General Metrics

This section is devoted to measuring and comparing the performance of staircase/K-norm for general ℓ_p metrics which do not necessarily conform to the sensitivity space K. Figure 6 and 8 illustrates different mechanisms' errors, respectively for when K is either one of the $\{\ell_p\}_{p=1,2,\infty}$ unit balls, or the ℓ_0-bounded "sum" shape as described in Section 5. Here we conduct the same experiments, measured in all ℓ_p-norms for $p \in \{1, 2, \infty\}$, in more scenarios with different (ε, d) values. In each of these scenarios, the γ for staircase mechanism is always chosen to be the one that minimizes the error when using $\|\cdot\|_K$ as the metric. Note that this γ does not necessarily minimize the ℓ_p-norm error that does not match K. However, we are able to show that the staircase mechanism always outperforms the vanilla K-norm, and, in the case of the ℓ_0-bounded "sum" polytope, the Laplace mechanism. As before, the Laplace mechanism is a special case of K-norm developed using the minimal ℓ_1-norm ball that contains the sensitivity space K. See Figure 13 and 14 for the results.

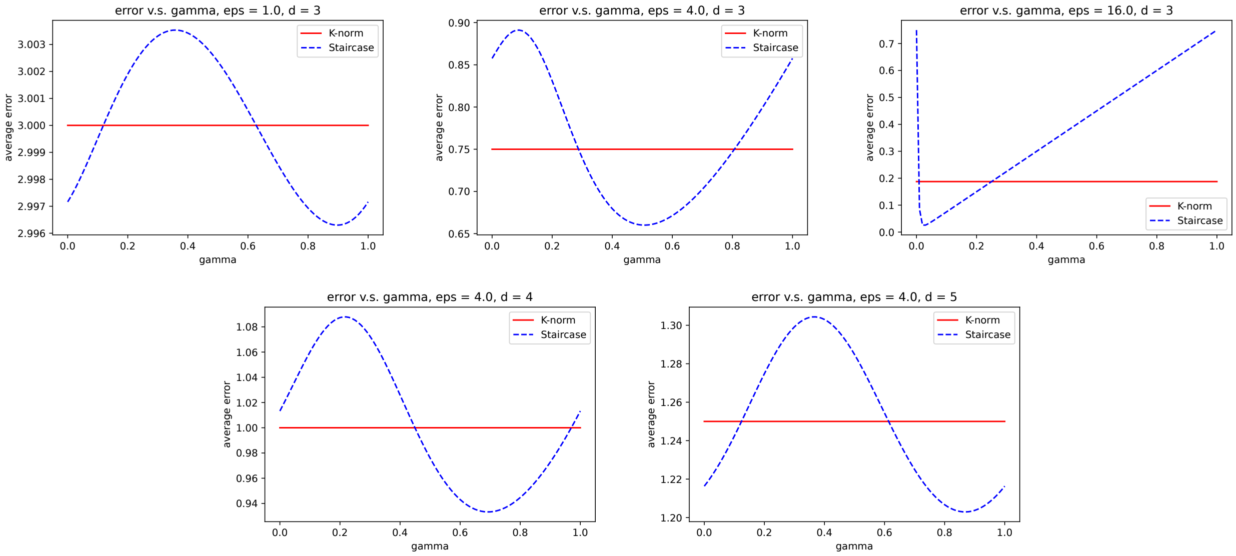

Figure 12: Error measured in $\|\cdot\|_K$ v.s. different γ values, in different (ε, d) scenarios

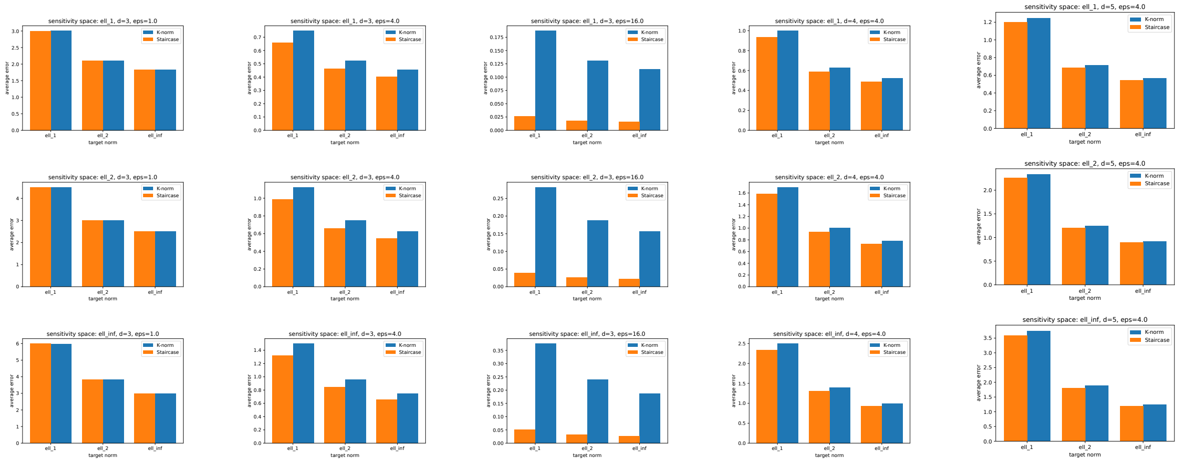

Figure 13: Expected error measured in ℓ_p-norms ($p = 1, 2, \infty$) of staircase/K-norm mechanisms with different ℓ_p-norm balls as sensitivity spaces, in (ε, d) scenarios

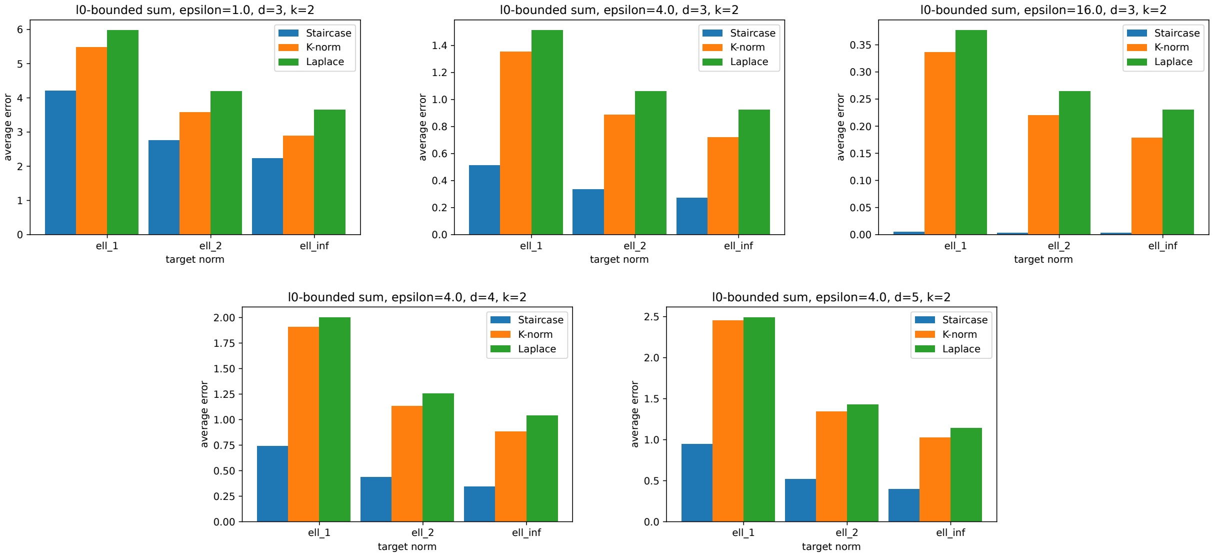
l0-bounded sum, epsilon=1.0, d=3, k=2
l0-bounded sum, epsilon=4.0, d=3, k=2
l0-bounded sum, epsilon=16.0, d=3, k=2
l0-bounded sum, epsilon=4.0, d=4, k=2
l0-bounded sum, epsilon=4.0, d=5, k=2
Staircase
K-norm
Laplace
average error
target norm
ell_1
ell_2
ell_inf
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
0.0
0.5
1.0
1.5
2.0
2.5

Figure 14