Better Gaussian Mechanism using Correlated Noise

Christian Janos Lebeda
IT University of Copenhagen

Problem formulation

Given a dataset $X = (x_1, \ldots, x_n)$ of n elements $x_i \in [0,1]^d$, design a private mechanism to approximate the sum:

$$f(X) := \sum_{i=1}^{n} x_i$$

For neighboring datasets $X \sim X'$ under add/remove one of these cases hold:
$|X'| = |X| + 1$ and $f(X') - f(X) \in [0,1]^d$
$|X'| = |X| - 1$ and $f(X') - f(X) \in [-1,0]^d$

We focus on variants of the Gaussian mechanism. We ignore scaling from privacy parameters for this poster.

The algorithm is simple!

The standard Gaussian mechanism adds i.i.d. noise from $\mathcal{N}(0,\Delta_2^2) = \mathcal{N}(0,d)$.

We add the same sample to all entries.

This allows us to reduce the i.i.d. noise.

Algorithm 1: Gaussian Mechanism using Correlated Noise

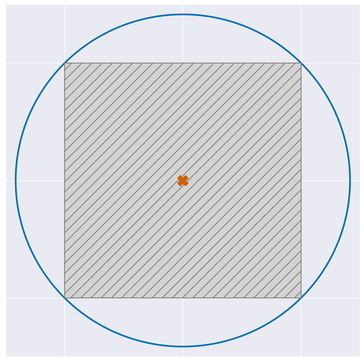
Input : Dataset $X = (x_1, x_2, \ldots, x_n)$ where $x_i \in [0,1]^d$.
Output : DP estimates of $f(X) := \sum_{i=1}^{n} x_i$ and n under add/remove.

Sample $\eta \sim \mathcal{N}(0, \frac{\sqrt{d}+1}{4})$.
foreach $i \in [d]$ **do**
  Sample $z_i \sim \mathcal{N}(0, \frac{d+\sqrt{d}}{4})$.
  Let $\tilde{x}_i \leftarrow f(X)_i + \eta + z_i$.
return $\tilde{x}, n + 2\eta$.

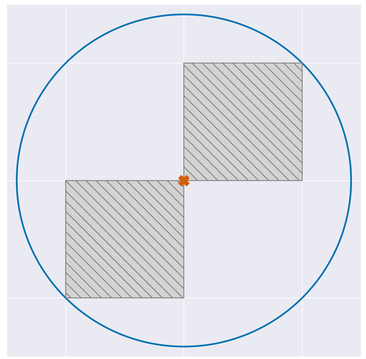
Geometric intuition

The standard Gaussian mechanism works well for replacement DP, but the noise does not fit the shape of data well under add/remove DP.

Standard Gaussian Mechanism



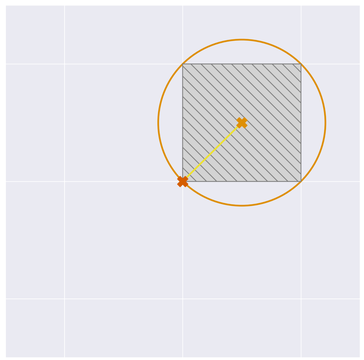
Replacement



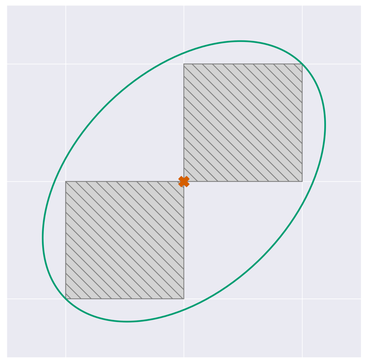
Add/remove

Our noise is elliptical around $(1,1,\ldots,1,1)$.

Correlated Gaussian Mechanism



"Add-DP ideal noise"



Add/remove

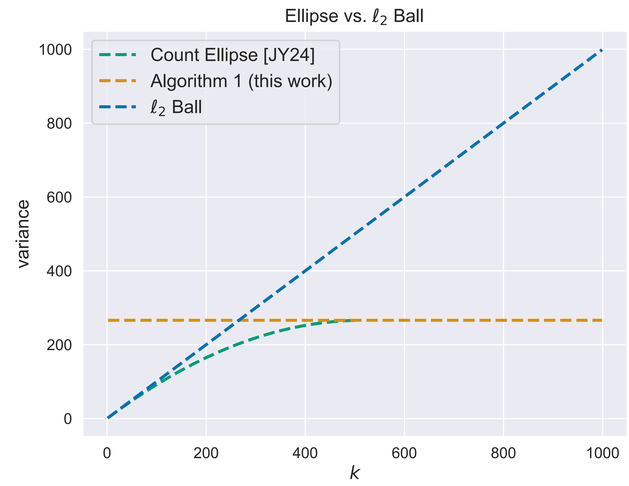
Main result

Error scales as $\left(\sqrt{d}+1\right)/2$ instead of $\sqrt{d}$.

Bounded contribution

[JY24] considered the case where users contribute at most k values ($\|x_i\|_0 \leq k$). Check out their work in Poster Session 2!

They give the optimal Gaussian noise mechanism for any $k \leq d/2$.



Error is (almost) optimal for $k \geq \lfloor d/2 \rfloor$.

Potential Future Work

- How can correlated noise improve error for thresholding techniques [WKZK24]?
- Can we improve error for hierarchical data with this technique? Other queries?

References

[JY24] Matthew Joseph and Alexander Yu. Some Constructions of Private, Efficient, and Optimal K-Norm and Elliptic Gaussian Noise.
[WKZK24] Arjun Wilkins, Daniel Kifer, Danfeng Zhang, and Brian Karrer. Exact Privacy Analysis of the Gaussian Sparse Histogram Mechanism.

arXiv

christian.j.lebeda@gmail.com