The Geometry of Differential Privacy: the Approximate and Sparse Cases

Aleksandar Nikolov[1] Kunal Talwar[2] Li Zhang[2]

Rutgers U.

Microsoft Research, SVC

Outline

Example

ID	Gender	Zip Code	Smoker	Lung Cancer
089341	M	07306	No	No
908734	F	10001	Yes	Yes
560671	M	08541	Yes	No

The data is both sensitive (medical information) and personally identifiable (with the right kind of side information).

Universe: All possible settings of the attributes

Histogram: Number of users for each setting of the attributes.

Queries:

- How many male smokers have lung cancer?
- How many more female smokers are there than male smokers?

Setting

- A *universe* U of user types; $|U| = N$
- A *database* $D \in U^n$ of n users, each having some type in U
- The database in *histrogram representation*:
  - $x \in \mathbb{R}^U$: x_i is the number of users in the database having type $i \in U$
  - $\|x\|_1 = \sum_{i \in U} |x_i| = n$
  - $D \triangle D' \leq 1 \Leftrightarrow \|x - x'\|_1 \leq 1$

Linear Queries

A useful and rich primitive: linear queries on the histrogram x.

- **Linear Query**: $\langle a, x \rangle$
- **Query Matrix**: d linear queries: Ax where $A \in \mathbb{R}^{d \times N}$
  - when A is a 0-1 matrix, we call the d queries *counting queries*

Privacy

Privacy Goal: compute *aggregate* statistics (here: linear queries) without revealing the type of any user, even to an adversary who knows the types of all other users.

Privacy

Privacy Goal: compute *aggregate* statistics (here: linear queries) without revealing the type of any user, even to an adversary who knows the types of all other users.

Definition

An algorithm $\mathcal{M}$ with input domain $\mathbb{R}^N$ and output range Y is (ε, δ)-*differentially private* if for every n, every x, x' with $\|x - x'\|_1 \leq 1$, and every measurable $S \subseteq Y$, $\mathcal{M}$ satisfies

$$\Pr[\mathcal{M}(x) \in S] \leq e^{\varepsilon} \Pr[\mathcal{M}(x') \in S] + \delta.$$

Privacy

Privacy Goal: compute *aggregate* statistics (here: linear queries) without revealing the type of any user, even to an adversary who knows the types of all other users.

Definition

An algorithm $\mathcal{M}$ with input domain $\mathbb{R}^N$ and output range Y is (ε, δ)-*differentially private* if for every n, every x, x' with $\|x - x'\|_1 \leq 1$, and every measurable $S \subseteq Y$, $\mathcal{M}$ satisfies

$$\Pr[\mathcal{M}(x) \in S] \leq e^{\varepsilon} \Pr[\mathcal{M}(x') \in S] + \delta.$$

Intuition: Algorithm does almost the same, no matter if a particular user participated or not. *Incentive to participate in a study*.

Accuracy

Accuracy of algorithm $\mathcal{M}$ – *mean squared error*:

$$\mathrm{Err}(\mathcal{M}, A, n) = \max_{x:\|x\|_1 \le n} \mathbb{E} \frac{1}{d} \|\mathcal{M}(A, x, n) - Ax\|_2^2$$

$$\mathrm{Err}(\mathcal{M}, A) = \max_n \mathrm{Err}(\mathcal{M}, A, n)$$

Accuracy

Accuracy of algorithm $\mathcal{M}$ – *mean squared error*:

$$\mathrm{Err}(\mathcal{M}, A, n) = \max_{x:\|x\|_1 \leq n} \mathbb{E} \frac{1}{d} \|\mathcal{M}(A, x, n) - Ax\|_2^2$$

$$\mathrm{Err}(\mathcal{M}, A) = \max_n \mathrm{Err}(\mathcal{M}, A, n)$$

Optimal error on A and on databases of size up to n is:

$$\mathrm{Opt}_{\varepsilon,\delta}(A, n) = \min_{\mathcal{M}} \mathrm{Err}(\mathcal{M}, A, n),$$

where the minimum is over all (ε, δ)-differentially private algorithms $\mathcal{M}$.

Accuracy

Accuracy of algorithm $\mathcal{M}$ – *mean squared error*:

$$\mathrm{Err}(\mathcal{M}, A, n) = \max_{x:\|x\|_1 \le n} \mathbb{E}\frac{1}{d}\|\mathcal{M}(A, x, n) - Ax\|_2^2$$

$$\mathrm{Err}(\mathcal{M}, A) = \max_n \mathrm{Err}(\mathcal{M}, A, n)$$

Optimal error on A and on databases of size up to n is:

$$\mathrm{Opt}_{\varepsilon,\delta}(A, n) = \min_{\mathcal{M}} \mathrm{Err}(\mathcal{M}, A, n),$$

where the minimum is over all (ε, δ)-differentially private algorithms $\mathcal{M}$.

The optimum when database size is unrestricted:

$$\mathrm{Opt}_{\varepsilon,\delta}(A) = \max_n \mathrm{Opt}_{\varepsilon,\delta}(A, n)$$

Universal bounds on error

For $A \in [0,1]^{d \times N}$:

- $\mathrm{Opt}_{\varepsilon,\delta}(A) = O(d)$
  - [DKM$^+$06]: Add $N(0, \sqrt{d}c(\varepsilon,\delta))$ noise to each query answer
  - [DN03]: Tight for random A

Universal bounds on error

For $A \in [0,1]^{d \times N}$:

- $\mathrm{Opt}_{\varepsilon,\delta}(A) = O(d)$
  - [DKM$^+$06]: Add $N(0, \sqrt{d}c(\varepsilon,\delta))$ noise to each query answer
  - [DN03]: Tight for random A
- $\mathrm{Opt}_{\varepsilon,\delta}(A, n) = O(n\sqrt{\log N})$
  - [HR10, GRU12]: Multiplicative weights, median mechanism
  - [DN03]: Tight up to the $\sqrt{\log N}$ for random A

Universal bounds on error

For $A \in [0,1]^{d \times N}$:

- $\mathrm{Opt}_{\varepsilon,\delta}(A) = O(d)$
  - [DKM+06]: Add $N(0, \sqrt{d}c(\varepsilon,\delta))$ noise to each query answer
  - [DN03]: Tight for random A
- $\mathrm{Opt}_{\varepsilon,\delta}(A, n) = O(n\sqrt{\log N})$
  - [HR10, GRU12]: Multiplicative weights, median mechanism
  - [DN03]: Tight up to the $\sqrt{\log N}$ for random A
- $\mathrm{Opt}_{\varepsilon,0}(A, n) = O(n^{4/3}\,\mathrm{polylog}(N))$
  - [BLR08]: Learning theoretic techniques
  - *This work*: $\mathrm{Opt}_{\varepsilon,0}(A, n) = O(n\,\mathrm{polylog}(N, d))$

Special A?

Some matrices A require a lot less error:

$$A = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}$$

Special A?

Some matrices A require a lot less error:

$$A = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}$$

- $\mathrm{Opt}_{\varepsilon,\delta}(A) = O(\mathrm{polylog}(d))$
  - Algorithm: answer a different set of queries, based on a binary tree data structure
  - Notice: A is TUM

Results

- An algorithm $\mathcal{M}$ is *α-optimal in the dense case* if it is (ε, δ)-d.p. and

$$\mathrm{Err}(\mathcal{M}, A) \le \alpha \, \mathrm{Opt}_{\varepsilon,\delta}(A)$$

- An algorithm $\mathcal{M}$ is *α-optimal in the sparse case* if it is (ε, δ)-d.p. and

$$\mathrm{Err}(\mathcal{M}, n) \le \alpha \, \mathrm{Opt}_{\varepsilon,\delta}(A, n)$$

$\alpha =$	**Unbounded** n (Dense)	**Bounded** n (Sparse)
$(\varepsilon, 0)$-d.p.	$\mathrm{polylog}(d)$ [1]	?
(ε, δ)-d.p.	?	?

Table : Values for α

[1] [HT10, BDKT12]

Results

- An algorithm $\mathcal{M}$ is *α-optimal in the dense case* if it is (ε, δ)-d.p. and

$$\mathrm{Err}(\mathcal{M}, A) \le \alpha \, \mathrm{Opt}_{\varepsilon,\delta}(A)$$

- An algorithm $\mathcal{M}$ is *α-optimal in the sparse case* if it is (ε, δ)-d.p. and

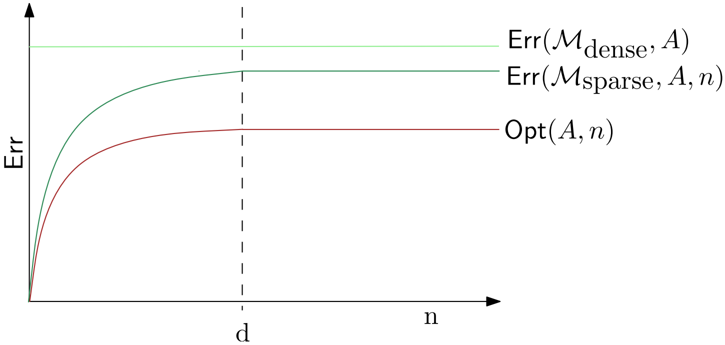
$$\mathrm{Err}(\mathcal{M}, n) \le \alpha \, \mathrm{Opt}_{\varepsilon,\delta}(A, n)$$

$\alpha =$	**Unbounded** n (Dense)	**Bounded** n (Sparse)
$(\varepsilon, 0)$-d.p.	$\mathrm{polylog}(d)$ [1]	$\mathrm{polylog}(d, N)$
(ε, δ)-d.p.	$\mathrm{polylog}(d)$	$\mathrm{polylog}(d, N)$

Table : Values for α

[1] [HT10, BDKT12]

Growth of Error with n



What the algorithms look like?

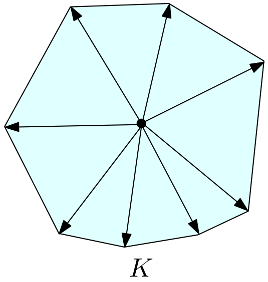
- *Dense case* ($n = \Omega(d)$)
  - Add correlated Gaussian noise w and output $\tilde{y} = Ax + w$
- *Sparse case* ($n = o(d)$)
  - Compute noisy answers $\tilde{y}$ using the dense case algorithm
  - Find the closest set of answers $\hat{y}$ that can be generated by a database x of size $\|x\|_1 \leq n$

Outline

The Lead Actor: K

Let $K = AB_1$ where B_1 is the ℓ_1 ball:

- nK is all query answers that can be generated by a size n-database.
- $K = \text{conv}\{\pm a_1, \dots, \pm a_N\}$



Preliminaries: Gaussian Mechanism

Basic algorithm $\mathcal{M}_{\mathrm{GN}}$:

- Say $K \subseteq rB_2^d$ (ℓ_2-sinsitivity is r)
- Output $Ax + w$, where $w \sim N(0, r \cdot c(\varepsilon, \delta))^d$

Properties:

- satisfies (ε, δ)-differential privacy
- $\mathrm{Err}(\mathcal{M}_{\mathrm{GN}}, A) = O(r^2)$

Preliminaries: Noise Lower Bounds

- [HT10]: $\mathrm{Opt}_{\varepsilon,0}(A) \geq d^2 \,\mathrm{vol}(K)^{2/d}$

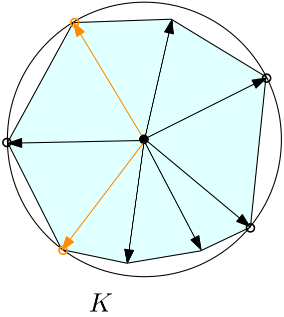
Preliminaries: Noise Lower Bounds

- [HT10]: $\mathrm{Opt}_{\varepsilon,0}(A) \geq d^2 \,\mathrm{vol}(K)^{2/d}$
- [MN12]: Say S is a simplex of d vertices of K and the origin
  $\Rightarrow$ $\mathrm{Opt}_{\varepsilon,\delta}(A) \geq d^2 \,\mathrm{vol}(S)^{2/d}$
  - lower bound uses combinatorial discrepancy

Preliminaries: Noise Lower Bounds

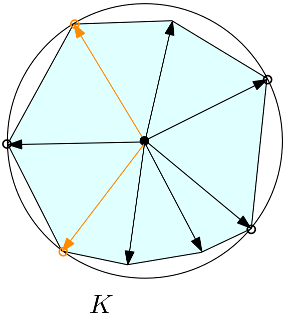
- [HT10]: $\mathrm{Opt}_{\varepsilon,0}(A) \geq d^2 \,\mathrm{vol}(K)^{2/d}$
- [MN12]: Say S is a simplex of d vertices of K and the origin $\Rightarrow$ $\mathrm{Opt}_{\varepsilon,\delta}(A) \geq d^2 \,\mathrm{vol}(S)^{2/d}$
  - lower bound uses combinatorial discrepancy
- $\mathrm{Opt}_{\varepsilon,\delta}(\Pi A) \leq \mathrm{Opt}_{\varepsilon,\delta}(A)$ for any projection Π $\Rightarrow$ can use lower bound on any ΠA to lower bound $\mathrm{Opt}_{\varepsilon,\delta}(A)$.

Preliminaries: The Löwner Ellipsoid



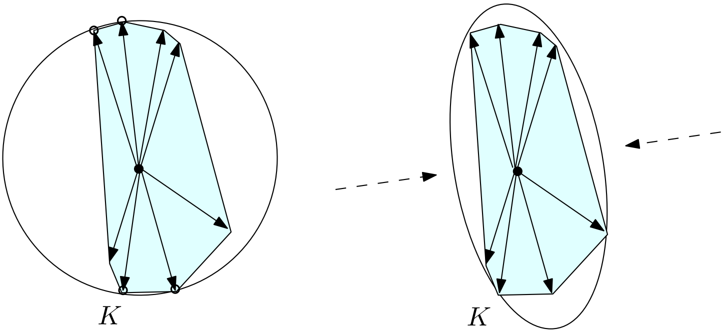
- Every K has a a unique minimum volume ellipsoid (MEE) containing it. [Joh48].
- [BT87, Ver01]: If the MEE of K is a ball rB_2^d, there are $\Omega(d)$ contact points of rB_2^d and K which are *pairwise nearly orthogonal*.

Preliminaries: The Löwner Ellipsoid



- Every K has a a unique minimum volume ellipsoid (MEE) containing it. [Joh48].
- [BT87, Ver01]: If the MEE of K is a ball rB_2^d, there are $\Omega(d)$ contact points of rB_2^d and K which are *pairwise nearly orthogonal*.

Preliminaries: The Löwner Ellipsoid



- Every K has a a unique minimum volume ellipsoid (MEE) containing it. [Joh48].
- [BT87, Ver01]: If the MEE of K is a ball rB_2^d, there are $\Omega(d)$ contact points of rB_2^d and K which are *pairwise nearly orthogonal*.

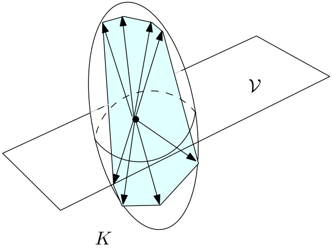
Optimality: pt 1

[BT87, Ver01]: If the MEE of K is a ball rB_2^d, there are $\Omega(d)$ contact points of rB_2^d and K which are *pairwise nearly orthogonal*.

- **When the MEE of K is a ball**:
  - Take the simplex S spanned by the nearly orthogononal contact points
  - $d^2 \operatorname{vol}(S)^{2/d} = \Omega(r^2)$
  - *The Gaussian Mechanism $Ax + N(0, r \cdot c(\varepsilon, \delta))$ is optimal!*

Optimality: pt 2

- **But when the MEE is a "long" ellipse?**:



  - Find a subspace $\mathcal{V}$ (of dimension $\Omega(d)$) such that $\Pi_{\mathcal{V}} E$ is like a sphere
  - Run Gaussian Mechanism on $\Pi_{\mathcal{V}} K$ and recurse on $\mathcal{V}^{\perp}$
  - Can still get a large simplex even inside $\mathcal{V}$ using the full power of [Ver01].

Outline

Sparse case noise lower bound

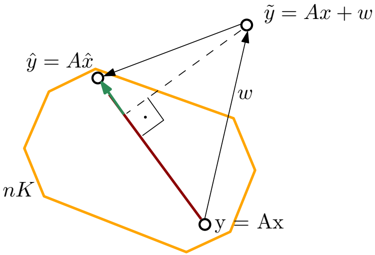
- If S is a simplex of $k \leq n$ vertices of K and the origin
  $\Rightarrow \ \mathrm{Opt}_{\varepsilon,\delta}(A, n) \geq \frac{1}{d} k^3 \,\mathrm{vol}(S)^{2/k}$
- *Notice*: when the MEE of K is a ball, we found a simplex S which is almost regular
  - $\Rightarrow$ *any face* of S gives a lower bound of $\Omega(\frac{n}{d} r^2)$
- But what algorithm matches the bound?

Simple Algorithm for Sparse Case

Gaussian Noise + Least Squares Estimation $\mathcal{M}_{\mathsf{GN\,+\,LSE}}$:

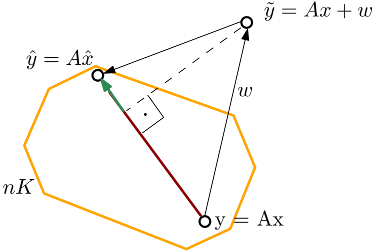
1. *Add noise*: Compute $\tilde{y} = Ax + w$ for $w \sim N(0, r \cdot c(\varepsilon, \delta))^d$
2. *Project*: Output $\arg\min\{\|\hat{y} - \tilde{y}\|_2 : \hat{y} \in nK\}$.

Optimality: MEE is a ball



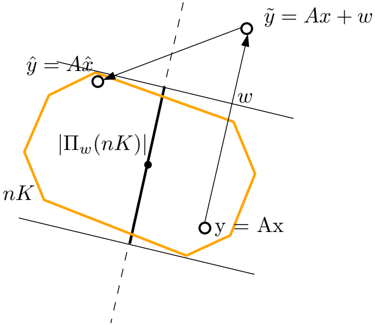
- $\frac{1}{d}\|\hat{y} - y\|_2^2 \leq \frac{4}{d}\|w\|_2^2$.

Optimality: MEE is a ball



- $\frac{1}{d}\|\hat{y} - y\|_2^2 \leq \frac{4}{d}\|w\|_2^2$.
- $\frac{1}{d}\|\hat{y} - y\|_2^2 \leq \frac{2}{d}|\langle w, \hat{y} - y\rangle|$.

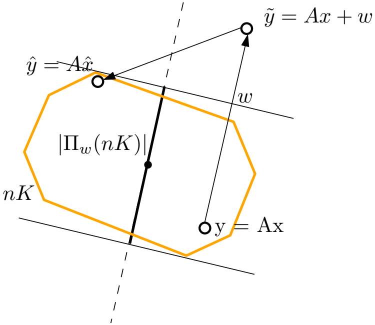
Optimality: MEE is a ball



- $\frac{1}{d}\|\hat{y} - y\|_2^2 \leq \frac{4}{d}\|w\|_2^2$.
- $\frac{1}{d}\|\hat{y} - y\|_2^2 \leq \frac{2}{d}|\langle w, \hat{y} - y\rangle|$.

$$
\begin{aligned}
\mathbb{E}\frac{2}{d}|\langle w, \hat{y} - y\rangle| &\leq \mathbb{E}\frac{4n}{d}\|A^T w\|_\infty \\
&= \mathbb{E}\frac{4}{d}|\Pi_w(nK)| \\
&\leq \frac{4n}{d} r^2 \sqrt{\log N}
\end{aligned}
$$

Optimality: MEE is a ball



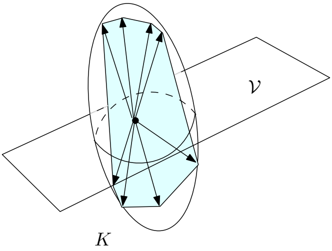
- $\frac{1}{d}\|\hat{y} - y\|_2^2 \leq \frac{4}{d}\|w\|_2^2$.
- $\frac{1}{d}\|\hat{y} - y\|_2^2 \leq \frac{2}{d}|\langle w, \hat{y} - y\rangle|$.

$$\begin{aligned}\mathbb{E}\frac{2}{d}|\langle w, \hat{y} - y\rangle| &\leq \mathbb{E}\frac{4n}{d}\|A^T w\|_\infty \\ &= \mathbb{E}\frac{4}{d}|\Pi_w(nK)| \\ &\leq \frac{4n}{d} r^2 \sqrt{\log N}\end{aligned}$$

- *The Gaussian Mechanism + LSE is nearly optimal!*

Optimality: General

Same ideas as before:



- Find a subspace $\mathcal{V}$ such that $\Pi_{\mathcal{V}} E$ is like a sphere
- Run Gaussian Mechanism + LSE on $\Pi_{\mathcal{V}} K$ and recurse on $\mathcal{V}^{\perp}$
- Full power of [Ver01] gives a lower bound.

Miscellanea

- $(\varepsilon, 0)$-differential privacy: use generalized K-norm noise of [HT10, BDKT12] to "approximate" Gaussian noise.
- In the dense case can extend to worst-case error per query using boosting
- Our lower bounds are in terms of hereditary discrepancy and our upper bounds are efficiently computable and nearly matching: *first polylogarithmic approximation to hereditary discrepancy*.

Summary and open questions

- A simple (ε, δ)-d.p. algorithm for answering linear queries optimally for any workload A and database size n.
- Improved on the error bound of [BLR08]
- Polylogarithmic approximation for hereditary discrepancy.

Questions:

- Can an algorithm that processes queries online be competitive?
- Other cases where simple least squares regression provably helps?
- Other data parameters that help reduce error?

Thank you!

Aditya Bhaskara, Daniel Dadush, Ravishankar Krishnaswamy, and Kunal Talwar.
Unconditional differentially private mechanisms for linear queries.
In *Proceedings of the 44th symposium on Theory of Computing*, STOC '12, pages 1269–1284, New York, NY, USA, 2012. ACM.

Avrim Blum, Katrina Ligett, and Aaron Roth.
A learning theory approach to non-interactive database privacy.
In *STOC '08: Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 609–618, New York, NY, USA, 2008. ACM.

J. Bourgain and L. Tzafriri.
Invertibility of large submatrices with applications to the geometry of banach spaces and harmonic analysis.
Israel journal of mathematics, 57(2):137–224, 1987.

Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor.
Our data, ourselves: Privacy via distributed noise generation, 2006.

Irit Dinur and Kobbi Nissim.
Revealing information while preserving privacy.
pages 202–210, 2003.

Anupam Gupta, Aaron Roth, and Jonathan Ullman.
Iterative constructions and private data release.
In *TCC*, pages 339–356, 2012.

M. Hardt and G. Rothblum.
A multiplicative weights mechanism for privacy-preserving data analysis.
Proc. 51st Foundations of Computer Science (FOCS). IEEE, 2010.

Moritz Hardt and Kunal Talwar.
On the geometry of differential privacy.
In *Proceedings of the 42nd ACM symposium on Theory of computing*, STOC '10, pages 705–714, New York, NY, USA, 2010. ACM.

F. John.
Extremum problems with inequalities as subsidiary conditions.

In *Studies and Essays presented to R. Courant on his 60th Birthday*, pages 187–204, 1948.

S. Muthukrishnan and Aleksandar Nikolov.
Optimal private halfspace counting via discrepancy.
In *Proceedings of the 44th symposium on Theory of Computing*, STOC '12, pages 1285–1292, New York, NY, USA, 2012. ACM.

R. Vershynin.
John's decompositions: Selecting a large part.
Israel Journal of Mathematics, 122(1):253–277, 2001.