Subspace Differential Privacy

Jie Gao, Ruobin Gong, Fang-Yi Yu
Data collection and release

• Examples
  – 2020 Census data by U.S. Census Bureau
  – Personal data in iOS or Chrome
  – Survey

• utility and privacy
Utility of data release

• Error magnitude
  – Mean-squared loss, zero-one loss

• Transparency and interpretability
  – Statistical inference

• External invariant constraint
  – For census data
    • population totals
    • counts of total housing units
    • group quarter and facilities
Utility and differential privacy
Utility and differential privacy

- Utility
  - Interpretability
  - Invariants
    - e.g., linearity
- Induced subspace
- Differential privacy
  - e.g., accuracy
  - e.g., simplicity
Outline

• Setting and challenges
  – Linear invariants
  – Differential privacy and induced subspace differential privacy (ISDP)
• Two approaches for DP to ISDP
  – Projection
  – Extension
• Discussion
  – Optimality
  – Statistical Considerations and Implementation
Outline

• Setting and challenges
  – Linear invariants
  – Differential privacy and induced subspace differential privacy (ISDP)

• Two approaches for DP to ISDP
  – Projection
  – Extension

• Discussion
  – Optimality
  – Statistical Considerations and Implementation
Invariants and differential privacy

- Setting
  - histogram \( x \in \mathbb{N}^x \),
  - a counting query \( A: \mathbb{N}^x \rightarrow \mathbb{N}^d \),
  - random mechanism \( M: \mathbb{N}^x \rightarrow \mathbb{N}^d \)

- \( \epsilon \)-DP: for all adjacent histograms \( x \) and \( x' \) and outcome \( y \)
  \[
  \Pr[M(x) = y] \leq e^\epsilon \Pr[M(x') = y]
  \]

- Linear invariant with a linear function \( C: \mathbb{N}^d \rightarrow \mathbb{N}^{dc} \)
  \[
  CM(x) = CA(x), \forall x \in \mathbb{N}^x
  \]
DP and invariants are incompatible.

- Let the set of database be $\mathbb{N}^4$, $A$ be the histogram, and $C$ be the sum of the first two coordinate.
- Two adjacent databases $x = (1,2,3,4)$ and $x' = (2,2,3,4)$.
- If $M$ is invariant with $C$, then
  \[ \Pr[C M(x) = 3] = 1 \] but \[ \Pr[C M(x') = 3] = 0 \]

\[ M \text{ cannot be differentially private} \]
``Post-processing'' on DP for invariants

• A common method to impose invariants is via “post-processing” using optimization/distance minimization, e.g. Census TopDown (Abowd et al., 2019).

• Issues
  – Not differentially private anymore
  – Systematic bias and obscurity
Systematic bias of "post-processing"
Induced subspace differential privacy

Relax differential privacy for linear invariant

• Given \( M : \mathbb{N}^x \rightarrow \mathbb{N}^d \) and a linear function \( C : \mathbb{N}^d \rightarrow \mathbb{N}^{dc} \)

\[
M(x) = M_{\parallel}(x) + M_{\perp}(x)
\]

where \( M_{\parallel}(x) \in \text{row}(C) \) and \( M_{\perp}(x) \in \text{null}(C) = N \)

– Linear invariant \( C \) implies \( CM_{\parallel}(x) = CM(x) = CA(x) \) is fixed.
– Subspace DP asks \( M_{\perp}(x) = \Pi_N M(x) \) is differentially private
Induced subspace differential privacy

Given $\epsilon, \delta \geq 0$, a query $A : \mathcal{X}^* \rightarrow \mathbb{R}^n$ and a linear equality invariant $C : \mathbb{R}^n \rightarrow \mathbb{R}^{nc}$ with null space $\mathcal{N} := \{v \in \mathbb{R}^n : Cv = 0\}$, a mechanism $M : \mathcal{X}^* \rightarrow \mathbb{R}^n$ is $(\epsilon, \delta)$-induced subspace differentially private for query $A$ and an invariant $C$ if

1. $M$ is $\mathcal{N}$-subspace $(\epsilon, \delta)$-differentially private, i.e.

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

for all $x \sim x'$ and $S \subseteq \mathcal{V}$, and

2. $M$ satisfies the linear equality invariant $C$, i.e.

$$Pr[CM(x) = CA(x)] = 1.$$
Outline

• Setting and challenges
  – Linear invariants
  – Differential privacy and subspace differential privacy

• Two approaches for DP to ISDP
  – Projection
  – Extension

• Discussion
  – Optimality
  – Statistical Considerations and Implementation
Two approaches for DP to ISDP

Projection framework
- Converting an existing DP mechanism $M$ to ISDP
  $\mathcal{M}(x) := A(x) + \Pi_N(M(x) - A(x))$
- Project the noise into null space
- Projected Gaussian $A(x) + \Pi_N e$
  the variance of $e$ is of order $\Delta_2(A)$

Extension framework
- Choose a DP mechanism $\hat{M}$ for query
  $\Pi_N A(x)$
  $\mathcal{M}(x) := \Pi_R A(x) + \hat{M}(x)$
- Augmenting a smaller private query invariant-compatibly
- Extended Gaussian $A(x) + Q_N e$
  - $Q_N$ is a rotation matrix of $N$
  - the variance of $e$ is of order $\Delta_2(Q_N^T A)$
Outline

• Setting and challenges
  – Linear invariants
  – Differential privacy and subspace differential privacy

• Two approaches for DP to ISDP
  – Projection
  – Extension

• Discussion
  – Optimality
  – Statistical Considerations and Implementation
Discussion

• Optimality
  – optimal DP for query $\Pi_N A = \text{optimal ISDP for } A$ and invariant $C$
  – Optimal ISDP from the correlated Gaussian mechanism (Nikolov et al 13)

• Unbiasedness
  – Projected and extended Gaussian/Laplace mechanism are unbiased

• Transparency and statistical intelligibility
Future directions

• General invariants
  – Inequality
  – Discrete output space

• Trade off between utility and privacy