A numerical verification framework for differential privacy in estimation

Yunhai Han[1] and Sonia Martínez[2]

[1] Georgia Institute of Technology

[2] University of California, San Diego

American Control Conference, July 10th, 2022

Introduction

- Differential privacy(df) makes it hard to distinguish outputs of a mechanism produced by adjacent inputs, which can help preserve the privacy of shared data.

- It is difficult to verify the df properties of the proposed estimation mechanisms[1][2][3] since they take values on continuous spaces, requiring to check for an infinite set of inequalities.

- The numerical verification framework mitigates this problem by partitioning the continuous space into a suitably chosen finite set of collection and making the evaluation wrt this partition.

- We confirm the df properties of a novel W_2 MHE, while comparing its performance with alternative estimators in simulation.

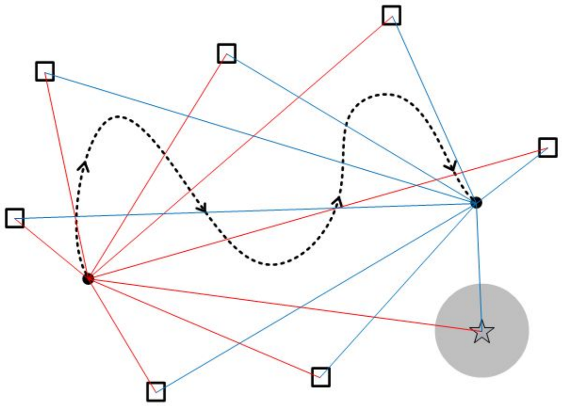


Figure 1. An example of differential privacy in sensor network

[1]. Cortes, G. E. Dullerud, S. Han, J. L. Ny, S. Mitra, and G. J. Pappas, "Differential privacy in control and network systems," in IEEE Int. Conf. on Decision and Control, 2016, pp. 4252–4272
[2]. E. Nozari, P. Tallapragada, and J. Cortes, "Differentially private distributed convex optimization via functional perturbation," IEEE Transactions on Control of Network Systems, 2019
[3]. J. L. Ny and G. J. Pappas, "Differentially private filtering," IEEE Transactions on Automatic Control, pp. 341–354, 2014

Problem Formulation

System & Observation model:

$$\Omega : \begin{cases} x_{k+1} = f\left(x_k, w_k\right), \\ y_k = h\left(x_k, v_k\right), \end{cases}$$

where $x_k \in \mathbb{R}^{d_X}$, $y_k \in \mathbb{R}^{d_Y}$, $w_k \in \mathbb{R}^{d_W}$ and $v_k \in \mathbb{R}^{d_V}$

A state estimator of this system is a stochastic mapping:

$$\mathbb{R}^{(T+1)d_Y} \to \mathbb{R}^{m d_X}, \text{ for some } m \geq 1$$

Differential privacy in estimation:

Definition 1 ((ε, d-Adjacent), λ-Approximate, Differential Privacy): Let $\mathcal{M}$ be a state estimator of System 1 and d_{y} a distance metric on $\mathbb{R}^{(T+1)d_Y}$. Given $\varepsilon, \lambda, d \in \mathbb{R}_{\geq 0}$, $\mathcal{M}$ is (ε, d-adjacent), λ-approximate, differentially private if for any $\mathrm{y}^1_{0:T}, \mathrm{y}^2_{0:T} \in \mathbb{R}^{(T+1)d_Y}$, with $d_{\mathrm{y}}(\mathrm{y}^1_{0:T}, \mathrm{y}^2_{0:T}) \leq d$ we have

$$\mathbb{P}(\mathcal{M}(\mathrm{y}^i_{0:T}) \in E) \leq e^{\varepsilon}\mathbb{P}\Big(\mathcal{M}\Big(\mathrm{y}^j_{0:T}\Big) \in E\Big)) + \lambda, \qquad (2)$$

- for $i, j = 1, 2$
- for all $E \subset \mathrm{range}(\mathcal{M})$
- $(\varepsilon, d\text{-adj})$, for $\lambda = 0$

Challenges & Solution

Technical challenges:

- Unknown range of the estimator -> High-likelihood differential privacy
- Infinite set of space partition -> Identification of a suitable space partition

High-likelihood differential privacy:

Definition 2: ***(High-likelihood** (ε,d**-adj**) **Differential Privacy).*** Suppose that $\mathcal{M}$ is a state estimator of System 1. Given $\varepsilon, d \in \mathbb{R}_{\geq 0}$, we say that $\mathcal{M}$ is (ε,d-adj) *differentially private with high likelihood* $1-\theta$ if there exists an event R with $\mathbb{P}(R) \geq 1-\theta$ such that, for any two $\mathrm{y}^i_{0:T}$, $i = 1, 2$, with $d_{\mathrm{y}}(\mathrm{y}^1_{0:T}, \mathrm{y}^2_{0:T}) \leq d$, we have:

$$\mathbb{P}(\mathcal{M}(\mathrm{y}^i_{0:T}) \in E | R) \leq e^{\varepsilon}\mathbb{P}(\mathcal{M}(\mathrm{y}^j_{0:T}) \in E | R),$$

for $i, j \in \{1, 2\}$ and all events $E \subseteq \mathrm{range}(\mathcal{M})$.

Lemma 1: Suppose that $\mathcal{M}$ is a high-likelihood (ε,d-adj) differentially private estimator, with likelihood $1-\theta$. Then, $\mathcal{M}$ is (ε,d-adj)-λ differentially private with $\lambda = \theta$.

Challenges & Solution

Identification of a suitable space partition:

Definition 3 ***(Differential privacy wrt a space partition):*** Let $\mathcal{M}$ be an estimator of System 1 and $\mathcal{P} = \{E_1, \ldots, E_n\}$ be a space partition[2] of range($\mathcal{M}$). We say that $\mathcal{M}$ is (ε,d-adj) differentially private wrt $\mathcal{P}$ if the definition of (ε,d-adj) differential privacy holds for each $E_k \in \mathcal{P}$.

Lemma 2: Let $\mathcal{M}$ be a state estimator of System 1, and consider a partition of range($\mathcal{M}$), $\mathcal{P}_1 = \{E_1, \ldots, E_{n_1}\}$, which is *finer* than another partition $\mathcal{P}_2 = \{F_1, \ldots, F_{n_2}\}$ ($n_1 > n_2$). That is, each F_i can be represented by the disjoint union $F_i = \cup_{s=1}^{m_i} E_{l_s}$. Then, if $\mathcal{M}$ is (ε,d-adj) differentally private wrt $\mathcal{P}_1$, then it is also differentially private wrt $\mathcal{P}_2$.

[2]By partition we mean a collection of mutually exclusive and collectively exhaustive set of events wrt $\mathbb{P}$.

Challenges & Solution

Identification of a suitable space partition:

Lemma 3: Consider a partition $\mathcal{P} = \{E_i\}_{i \in \mathcal{I}}$ such that $\mathbb{P}(E_i) \leq \eta$ for all $i \in \mathcal{I}$. Then, if (ε,d-adj) differential privacy holds wrt the partition $\mathcal{P}$, then $\mathcal{M}$ is (ε,d-adj)-λ differentially private with $\lambda = 2\eta e^{\varepsilon}$.

The original problem is now turned into checking the differential privacy with respect to a high-likely range and a given partition of that range.

Test Framework

Overview:

Algorithm 1 (ε,d-adj) Differentially-private Test Framework

function TEST FRAMEWORK($\mathcal{M}, \varepsilon, \mathrm{y}^1_{0:T}, \mathrm{y}^2_{0:T}$)
  Inputs: Target estimator $\mathcal{M}$, privacy level ε, sensor data ($\mathrm{y}^1_{0:T}, \mathrm{y}^2_{0:T}$)
  `EventList` = `EventListGenerator`($\mathcal{M}, \mathrm{y}^1_{0:T}$)
  `WorstEvent` =
  `WorstEventSelector`($\mathcal{M}, \varepsilon, \mathrm{y}^1_{0:T}, \mathrm{y}^2_{0:T}$,
    `EventList`)
  p^+, p_+ = `HypothesisTest`($\mathcal{M}, \varepsilon, \mathrm{y}^1_{0:T}, \mathrm{y}^2_{0:T}$,
    `WorstEvent`)
  Return p^+, p_+
end function

* Test framework is inspired by the work: Z. Ding, Y. Wang, G. Wang, D. Zhang, and D. Kifer, “Detecting violations of differential privacy,” in Proc. ACM SIGSAC Conf. Comput. Commun. Security, 2018, pp. 475–489.

Test Framework

Event list generation:

Algorithm 2 `EventListGenerator`

function EVENTLISTGENERATOR($\mathcal{M}$, $y^1_{0:T}$, β, γ)
    Input: Target Estimator($\mathcal{M}$)
        Sensor Data($y^1_{0:T}$)
        Parameters for Algorithm 3 (β, γ)
    `HighLikelySet` ← Apply Algorithm 3
    `EventList` ← a partition of the `HighLikelySet`
    Return `EventList`
end function

Algorithm 3 `HighLikelySet`

Input: Target Estimator($\mathcal{M}$) with dimension d_X
    Sensor data($y^1_{0:T}$), parameters β, γ
Output: Matrix A^k and vector b^k representing an
    1-β-accurate high-likely set at time step k
    $R_k(A^k, b^k) = \{x \in \mathbb{R}^{d_X} \mid \|A^k x + b^k\|_2 \leq 1\}$
    with confidence $1 - \gamma$.
Set number of samples Γ =
    $\left\lceil \frac{1}{\beta} \frac{e}{e-1} \left(\log \frac{1}{\gamma} + d_X(d_X + 1)/2 + d_X \right) \right\rceil$
for $k \in \{0, \ldots, T\}$ **do**
    for $i \in \{0, \ldots, \Gamma\}$ **do**
        Record $z_i^k = \mathcal{M}\left(y^1_{0:k}\right)$
    end for
    Solve the convex problem
    $$\begin{aligned} &\arg\min_{A^k, b^k} && -\log\det A^k \\ &\text{subject to} && \left\|A^k z_i^k - b^k\right\|_2 - 1 \leq 0,\ i = 0, \ldots, \Gamma \end{aligned}$$
    return A^k, b^k
end for

* HighLikelySet method is inspired by the work: A. Devonport and M. Arcak, “Estimating reachable sets with scenario optimization,” in Proc. Annu. Learn. Dyn. Control Conf., 2020, pp. 75–84.

Test Framework

Hypothesis Test:

Algorithm 4 `WorstEvent` Selector

function WORSTEVENTSELECTOR(n, $\mathcal{M}$, ε, $y^1_{0:T}$, $y^2_{0:T}$, `EventList`)
  Input: Target Estimator($\mathcal{M}$)
    Desired differential privacy(ε)
    $d$adjacent sensor data($y^1_{0:T}$, $y^2_{0:T}$)
    `EventList`
  $O_1 \leftarrow$ Estimate set after n runs of $\mathcal{M}(y^1_{0:T})$
  $O_2 \leftarrow$ Estimate set after n runs of $\mathcal{M}(y^2_{0:T})$
  `pvalues` $\leftarrow$ [ ]
  for $E \in$ `EventList` **do**
    $c_1 \leftarrow |\{i | O_1[i] \in E\}|$
    $c_2 \leftarrow |\{i | O_2[i] \in E\}|$
    $p^+, p_+ \leftarrow$ PVALUE $(c_1, c_2, n, \varepsilon)$
    $p^* \leftarrow \min(p^+, p_+)$
    `pvalues`.append(p^*)
  end for
  `WorstEvent` $\leftarrow$ `EventList`[argmin{`pvalues`}]
  Return $E^* =$ `WorstEvent`
end function

Algorithm 5 `HypothesisTest`

function PVALUE(c_1, c_2, n, ε)
  $\bar{c_1} \leftarrow B(c_1, 1/e^{\varepsilon})$
  $s \leftarrow \bar{c_1} + c_2$
  $p^+ \leftarrow$ 1 - Hypergeom.cdf($\bar{c_1} - 1 | 2n, n, s$)
  $\bar{c_2} \leftarrow B(c_2, 1/e^{\varepsilon})$
  $s \leftarrow \bar{c_2} + c_1$
  $p_+ \leftarrow$ 1 - Hypergeom.cdf($\bar{c_2} - 1 | 2n, n, s$)
  return p^+, p_+
end function
function HYPOTHESISTEST(m, $\mathcal{M}$, ε, $y^1_{0:T}$, $y^2_{0:T}$, E^*)
  Input: Target Estimator($\mathcal{M}$)
    Desired differential privacy(ε)
    d-adjacent sensor data($y^1_{0:T}$, $y^2_{0:T}$)
    E^*(`WorstEvent`)
  $O_1 \leftarrow$ Estimate set after m runs of $\mathcal{M}(y^1_{0:T})$
  $O_2 \leftarrow$ Estimate set after m runs of $\mathcal{M}(y^2_{0:T})$
  $c_1 \leftarrow |\{i | O_1[i] \in E^*\}|$
  $c_2 \leftarrow |\{i | O_2[i] \in E^*\}|$
  $p^+, p_+ \leftarrow$ PVALUE $(c_1, c_2, m, \varepsilon)$
  Return p^+, p_+
end function

* Numerical test method is inspired by the work: R. A. Fisher, The Design of Experiments. Edinburgh, U.K.: Oliver Boyd, 1935, pp. 252–254.

Test Framework

Theoretical guarantee:

Theorem 1: Let $\mathcal{M}$ be a state estimator of System 1, and let ε, d, β and $\gamma \in \mathbb{R}_{\geq 0}$. We denote two d-adjacent sensor data as $\mathrm{y}^i_{0:T}$, $i \in \{1, 2\}$ and a partition of the high-likely $(1 - \beta)$ set R from Algorithm 3 with high confidence $1 - \gamma$ as $\mathcal{P} = \{E_1, \ldots, E_n\}$ such that $\mathbb{P}(E_i) \leq \eta$ for all i. Then, if Γ is selected accordingly, and the estimator passes the test in Algorithm 1, then $\mathcal{M}$ is approximately $(\varepsilon, d\text{-adj})$ differentially private wrt $\mathrm{y}^i_{0:T}$, $i \in \{1, 2\}$, and $\lambda = \beta + 2\eta e^{\varepsilon}$, with confidence $(1 - \alpha)(1 - \gamma)$.

Experiments

System & Observation model:

We consider a two-dimension non-isotropic model ($\mathbf{x}_k = (x_k^1, x_k^2)$) with the observation model as:

$$\begin{aligned} \mathrm{y}_k^i &= h(\mathbf{x}_k, \mathbf{q}_i) + \mathbf{v}_{i,k} \\ &= 100 \tanh(0.1(\mathbf{x}_k - \mathbf{q}_i)) + \mathrm{v}_k^i, \quad i = 1, \ldots, 10, \end{aligned}$$

where $\mathbf{q}_i \in \mathbb{R}^2$ is the position of sensor i .

Generate two d-adjacent sensor data: $\mathrm{y}_{0:T}^1, \mathrm{y}_{0:T}^2$

Implement the numerical framework on W_2-MHE filter

- Γ $(= 814)$
- $\beta = 0.05, \gamma = 10^{-9}$
- $T = 8, N = 5$
- $r = 2$
- s_k = 1, 0.8 or 0.7 (filter)

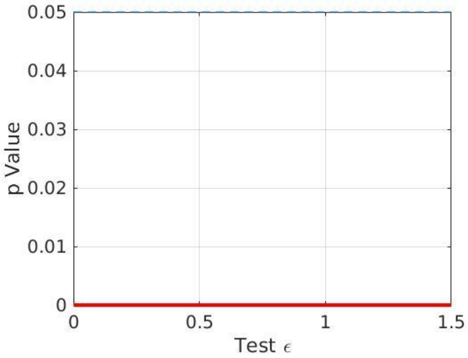
Compute →

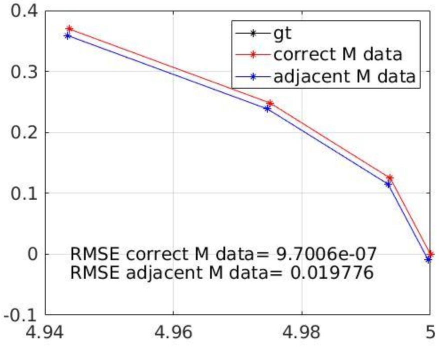
- Level of differential privacy ε
- Confidence value λ
- Estimation accuracy E_{correct}

Experiments

Test results

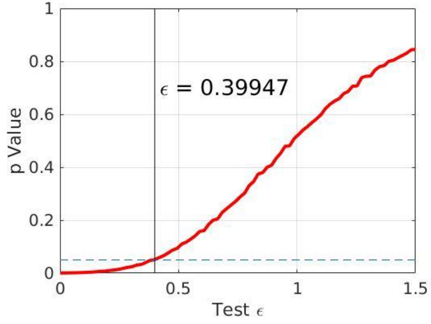
- $s_k = 1$

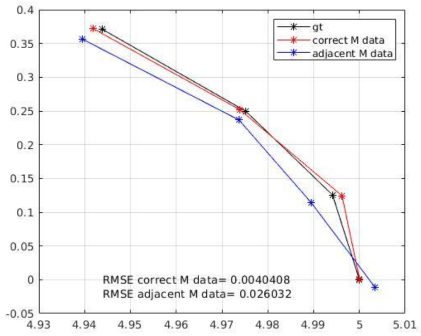




- Data distinguishable
- Very accurate (0 error)

- $s_k = 0.8$





- $\varepsilon = 0.39947$
- $\lambda = 0.0888$
- $E_{\text{correct}} = 0.004$

Experiments

Comparisons between different mechanisms

- W_2-MHE filter vs Input Perturbation

Sensor Setup	W_2MHE	Input Perturbation	Better choice
$\mathbf{Q_1}$	$\varepsilon_c = 0.39947$ $\lambda = 0.0888$ $E_{\text{correct}} = 0.0040408$	$\varepsilon_c = 0.41408$ λ=0.0803 $E_{\text{correct}} = 0.0013998$	Input Pert
$\mathbf{Q_2}$	$\varepsilon_c = 0.53229$ $\lambda = 0.1011$ $E_{\text{correct}} = 0.0049874$	$\varepsilon_c = 0.72204$ λ=0.2106 $E_{\text{correct}} = 0.0049674$	W_2-MHE
$\mathbf{Q_3}$	$\varepsilon_c = 0.98768$ $\lambda = 0.1037$ $E_{\text{correct}} = 0.0030866$	$\varepsilon_c = 2.3423$ $\lambda = 0.8408$ $E_{\text{correct}} = 0.0037826$	W_2-MHE

→

- Specific to sensor setup
- 2 out of 3, filter wins

- W_2-MHE filter vs *Differentially private EKF*

Sensor Setup	ε_c	E_{correct}	Better choice
$\mathbf{Q_1}$	0.46223	0.0066205	W_2-MHE
$\mathbf{Q_2}$	1.9239	0.0064686	W_2-MHE
$\mathbf{Q_3}$	2.3085	0.0062608	W_2-MHE

→

- W_2-MHE filter is better

* Differential private EKF is inspired by the work: J. L. Ny and G. J. Pappas, “Differentially private filtering,” IEEE Transactions on Automatic Control, pp. 341–354, 2014.

Conclusions

- A numerical test framework to evaluate the differential privacy of continuous-range mechanisms with a precise quantifiable performance guarantee

- A tool for the system designers to choose which differential-private mechanism to be used based on the numerical test results

Thank you for your time!

OFFICE OF NAVAL RESEARCH

Funding: ONR N00014-19-1-2471