Online Robust Locally Differentially Private Learning for Nonparametric Regression

Chenfei Gu[1], Qiangqiang Zhang[2], Ting Li[1], Jinhan Xie[3], Niansheng Tang[3]

[1] *School of Statistics and Data Science, Shanghai University of Finance and Economics*
[2] *Zhongtai Securities Institute for Financial Studies, Shandong University*
[3] *Yunnan Key Laboratory of Statistical Modeling and Data Analysis, Yunnan University*

Online Private Nonparametric Regression

Motivation

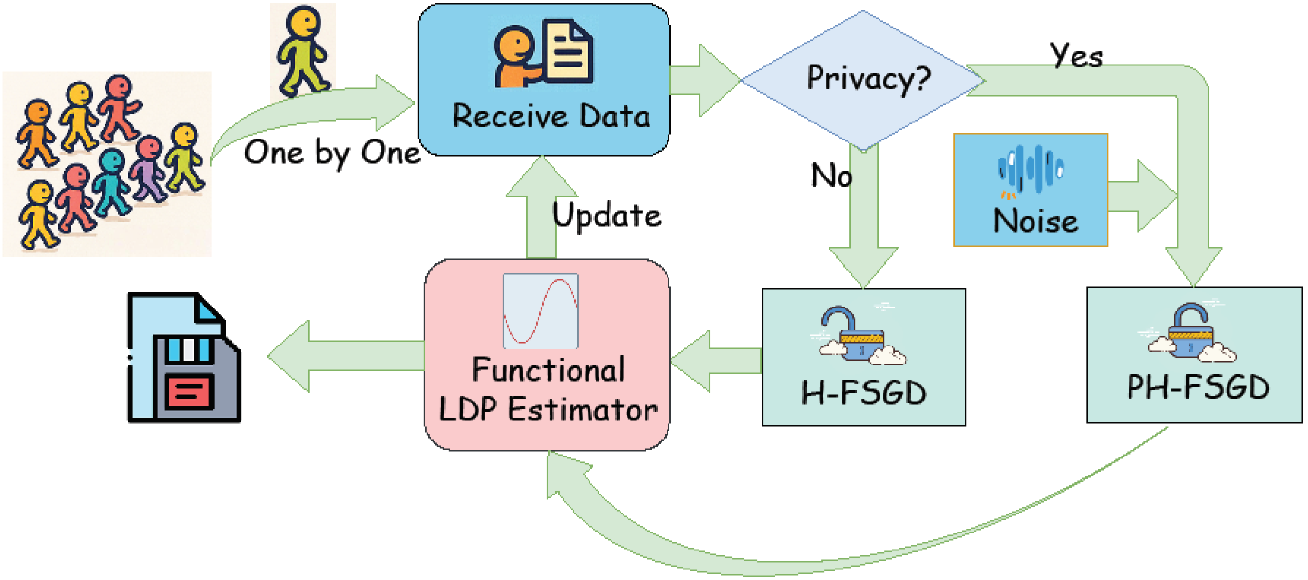
- Streaming data in autonomous systems, wearables, and healthcare require real-time, privacy-preserving models.
- Batch-based nonparametric methods are unsuitable for large-scale or continuous data.
- Outliers and heavy-tailed noise in streaming data hinder model reliability.
- Achieving local differential privacy (LDP) remains difficult beyond centralized or low-dimensional settings.

Table 1: A comparison of recent results on nonparametric regression.

Method	Online	One-pass	Robust	Optimal rate	Privacy
Hall et al. [2013]	✗	✗	✗	?	✓
Dieuleveut and Bach [2016]	✓	✓	✗	✓	✗
Liu et al. [2023]	✓	✓	✗	✓	✗
Quan and Lin [2024]	✓	✓	✗	✓	✗
Proposed	✓	✓	✓	✓	✓

Models and Problem Formulation

Goal



- Develop an online, private, nonparametric regression framework that is robust to heavy-tailed noise and satisfies LDP.
- Study non-asymptotic convergence guarantees and identify optimal step-size schedules.

Models and Problem Formulation

Setup

- The observed data are streaming samples $\{(X_n, Y_n)\}_{n=1}^{\infty}$ generated from the model $Y_n = f^{\star}(X_n) + e_n$.

- The best reproducing kernel Hilbert space (RKHS) approximation:

$$f_{\mathcal{H}} := \text{argmin}_{f \in \overline{\mathcal{H}}} \mathbb{E}\left[\left(Y - f(X)\right)^2\right].$$

- Our objective is to develop a computationally efficient, single-pass sequence of estimators for $f_{\mathcal{H}}$.

- To address robustness in the presence of heavy-tailed noises and to facilitate private updates, consider the Huber regression in an RKHS:

$$min_{f \in \mathcal{H}} \mathbb{E} L_{\tau}(Y - f(X)),$$

where $L_{\tau}(u) = \frac{1}{2} u^2 \mathbb{I}\{|u| \le \tau\} + (\tau|u| - \frac{1}{2}\tau^2)\mathbb{I}\{|u| > \tau\}$.

Methodology

Private Huber Functional SGD

Algorithm 1 PH-FSGD

Input: The streaming data $\{(X_n, Y_n)\}_{n\in\mathbb{N}}$, the initial estimates $\bar{f}(\cdot) = \hat{f}(\cdot) = 0$, the step size sequences $\{\gamma_n\}_{n\in\mathbb{N}}$, the tuning parameter $\tau > 0$, the reproducing kernel K, the bounded parameter $B > 0$, the privacy parameters $\{\varepsilon_n\}_{n\in\mathbb{N}}$, $\{\delta_n\}_{n\in\mathbb{N}}$, and the function grids $\{t_j\}_{j=1}^{J}$.

for $n = 1, 2, \ldots$ **do**

  Generate the noise $\{\xi_n(t_j)\}_{j=1}^{J}$ from $N_J(\mathbf{0}, \frac{8\tau^2 B^2 \log(2/\delta_n)}{\varepsilon_n^2} K^{(t)})$, where $K^{(t)}$ is a $J \times J$ matrix with its components $(K^{(t)})_{ij} = K(t_i, t_j)$.

  Calculate the residual: $\mathrm{res}_n = Y_n - \langle \hat{f}_{n-1}, K_{X_n} \rangle_{\mathcal{H}}$.

  Perform the noisy gradient descent at each function grid t_j for $j = 1, \ldots, J$ as follows.

    if $|\mathrm{res}_n| \le \tau$

      then $\hat{f}_n(t_j) = \hat{f}_{n-1}(t_j) + \gamma_n \mathrm{res}_n K(X_n, t_j) + \gamma_n \xi_n(t_j)$.

      elseif $\mathrm{res}_n > \tau$

        then $\hat{f}_n(t_j) = \hat{f}_{n-1}(t_j) + \gamma_n \tau K(X_n, t_j) + \gamma_n \xi_n(t_j)$.

      else $\hat{f}_n(t_j) = \hat{f}_{n-1}(t_j) - \gamma_n \tau K(X_n, t_j) + \gamma_n \xi_n(t_j)$.

  Update $\bar{f}_n$ at each function grid:

$$\bar{f}_n(t_j) = \frac{n-1}{n} \bar{f}_{n-1}(t_j) + \frac{1}{n} \hat{f}_n(t_j), j = 1, \ldots, J.$$

end for

Output: The estimators $\{\bar{f}_n(t_j)\}_{j=1}^{J}$ at each function grid t_j and each iteration n.

Theory

Constant Step Size Scheme

Table 3: Constant step size: optimal ζ and convergence rates.

r range	Optimal ζ in $\gamma_i \asymp n^{-\zeta}$	Private / non-private convergence rate
$(0, (\alpha-1)/(2\alpha)]$	0	$O(n^{-2r})$
$((\alpha-1)/(2\alpha), 1]$	$(2r\alpha+1-\alpha)/(2r\alpha+1)$	$O\big(n^{-2r\alpha/(2r\alpha+1)}\big)$
$(1, (\alpha+2)/2]$	$(\alpha+1)/(2r\alpha+1)$	$O\big(n^{-(2r\alpha-2r+2)/(2r\alpha+1)}\big)$
$((\alpha+2)/2, \infty)$	$1/(1+\alpha)$	$O\big(n^{-\alpha/(1+\alpha)}\big)$

Non-constant Step Size Scheme

Table 4: Non-constant step size: optimal ζ and convergence rates.

r range	Optimal ζ in $\gamma_i \asymp n^{-\zeta}$	Private / non-private convergence rate
$(0, (\alpha-1)/(2\alpha)]$	0	$O(n^{-2r})$
$((\alpha-1)/(2\alpha), (1+\alpha)/(2\alpha))$	$(2r\alpha+1-\alpha)/(2r\alpha+1+\alpha)$	$O\big(n^{-(2r\alpha+\alpha-1)/(2r\alpha+1+\alpha)}\big)$
$[(1+\alpha)/(2\alpha), \infty)$	$1/(1+\alpha)$	$O\big(n^{-\alpha/(1+\alpha)}\big)$

Experiment

Non-Private Synthetic Data

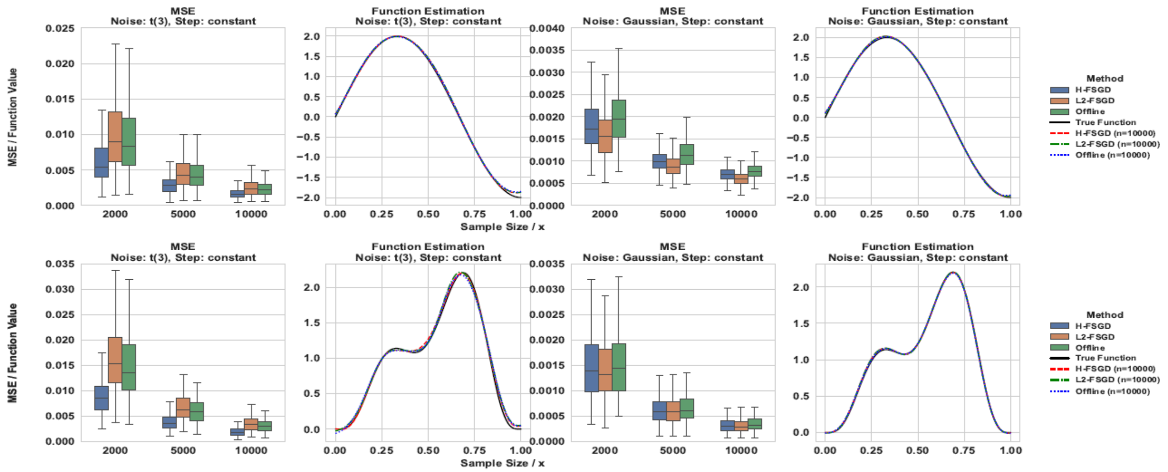


Figure 2: Box-plots and function fitting plots for Case 1 (top panels) and Case 2 (bottom panels) with the constant step size scheme in Example 5.1.

Private Synthetic Data

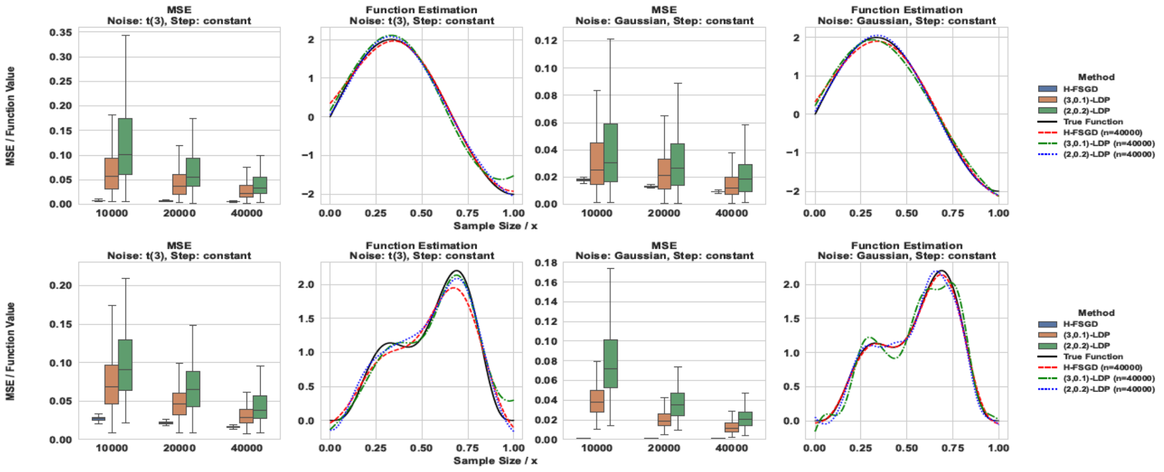


Figure 3: Box-plots and function fitting plots for Case 1 (top panels) and Case 2 (bottom panels) with the constant step size scheme in Example 5.2.

- The proposed H-FSGD method significantly outperforms the least-squares-based FSGD under heavy-tailed noises.
- PH-FSGD can still recover the true function shape well, even under strong privacy constraints.
- Stronger privacy enhances protection but also leads to greater estimation error and slower convergence.

Takeaway Notes

- **Online Robust LDP Estimation Framework**

  Develop an online robust LDP framework enabling per-iteration privacy guarantees and outlier-resistant real-time nonparametric regression in dynamic environments.

- **One-pass Algorithms**

  Propose two one-pass algorithms, H-FSGD and PH-FSGD, that achieve $O(1)$ time and space complexity per iteration without storing past observations.

- **Non-asymptotic Analysis**

  Establish comprehensive non-asymptotic convergence guarantees and identify optimal step-size schedules that achieve minimax-optimal rates.

Thank You!