# High Dimensional Sparse Linear Regression under Local Differential Privacy: Power and Limitations

**Di Wang**
Department of Computer Science and Engineering
State University of New York at Buffalo
Buffalo, NY, 14260

**Adam Smith**
Department of Computer Science
Boston University
Boston, MA, 02215

**Jinhui Xu**
Department of Computer Science and Engineering
State University of New York at Buffalo
Buffalo, NY, 14260

## Abstract

In this paper, we study high dimensional sparse linear regression under the Local Differential Privacy (LDP) model, and give both negative and positive results. On the negative side, we show that polynomial dependency on the dimensionality $p$ of the space is unavoidable in the estimation error under the non-interactive local model, if the privacy of the whole dataset needs to be preserved. Similar limitations also exist for other types of error measurements and in the (sequential) interactive local. This indicates that differential privacy in high dimensional space is unlikely achievable for the problem. On the positive side, we show that the optimal rate of the error estimation can be made logarithmically depending on $p$ (i.e., $\log p$) under the local model, if only the privacy of the responses (labels) is to be preserved, where the upper bound is obtained by a new method called Differentially Private Iterative Hard Thresholding (DP-IHT), which is interesting in its own right.

## 1 Introduction

In the paper, we study the locally differentially private version of the high dimensional sparse linear regression problem, where each user $i \in [n]$ holds a data record $(x_i, y_i) \in \mathbb{R}^p \times \mathbb{R}$. There are two commonly used ways for measuring the performance of this problem, which correspond to two different settings, the statistical learning and the statistical estimation settings. For the first setting, the measurement is based on the optimization error, *i.e.* $F(\theta^{\text{priv}}) - \min_{\theta \in C} F(\theta)$, where $F(\theta) = \mathbb{E}_{(x,y)\sim\mathcal{P}}(\langle x, w\rangle - y)^2$, and $\mathcal{P}$ is an unknown distribution. For the second setting, $y$ is assumed to be $y = \langle x, \theta^*\rangle + \sigma$, where $x \sim \mathcal{D}$, $\mathcal{D}$ is a known distribution, $\sigma$ is a random noise, and $\theta^* \in \mathbb{R}^p$ is the to-be-estimated vector that satisfies the condition of $\|\theta^*\|_0 = s$. The estimation error for this setting is represented by the loss of the squared $\ell_2$ norm, *i.e.,* $\|\theta^{priv} - \theta^*\|_2^2$. In this paper, we will focus on the latter setting, and assume that $x \sim \text{Uniform}\{+1, -1\}^p$.

Our contributions can be summarized as follows:

- We first present a negative result which suggests that the $\epsilon$ non-interactive private minimax risk (see Definition 2) of $\|\theta^{\text{priv}} - \theta^*\|_2^2$ is lower bounded by $\Omega(\frac{p \log p}{n\epsilon^2})$ if the privacy of the whole dataset $\{(x_i, y_i)\}_{i=1}^n$ needs to be preserved. This indicates that it is impossible to obtain any non-trivial error bound in high dimensional space (*i.e.* $p \gg n$). The private minimax risk is still lower bounded by $\Omega(\frac{p}{n\epsilon^2})$, even under the sequentially interactive local

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

model. Our proofs are based on a locally differentially private version of the Fano and Le Cam method [3, 4, 5]. We further reveal that this polynomial dependency on $p$ cannot be avoided even if we relax the measurement of the loss function.

- We then give a positive result for the case where only the responses (labels) are required to be private, *i.e.,* the dataset $\{x_i\}_{i=1}^n$ is assumed to be public and $\{y_i\}_{i=1}^n$ is private (note that this is a valid case as shown in [2, 1]). For this case, we propose a general algorithm called Differentially Private Iterative Hard Thresholding (DP-IHT), whose output can achieve an upper bound of $O(\frac{s\log p}{n\epsilon^2})$ for the estimation error. We show that this bound is actually optimal, as the $\epsilon$ non-interactive private minimax risk can also be lower bounded by $\Omega(\frac{s\log p}{n\epsilon^2})$, where $\Omega(\frac{s\log p}{n})$ is the optimal minimax rate of the non-private case [7]. As a general technique for differential privacy, DP-IHT is interesting in its own right, and can be potentially used to other problems.

## 2 Preliminaries

### 2.1 Classical Minimax Risk

Let $\mathcal{P}$ be a class of distributions over a data universe $\mathcal{X}$. For each distribution $p \in \mathcal{P}$, there is a deterministic function $\theta(p) \in \Theta$, where $\Theta$ is the parameter space. Let $\rho : \Theta \times \Theta :\mapsto \mathbb{R}_+$ be a semi-metric function on the space $\Theta$ and $\Phi : \mathbb{R}_+ \mapsto \mathbb{R}_+$ be a non-decreasing function with $\Phi(0) = 0$ (in this paper, we assume that $\rho(x, y) = |x-y|$ and $\Phi(x) = x^2$ unless specified otherwise). We further assume that $\{X_i\}_{i=1}^n$ are $n$ i.i.d observations drawn according to some distribution $p \in \mathcal{P}$, and $\hat{\theta} : \mathcal{X}^n \mapsto \Theta$ be some estimator. Then the minimax risk in metric $\Phi\circ\rho$ is defined by the following saddle point problem: $\mathcal{M}_n(\theta(\mathcal{P}), \Phi\circ\rho) := \inf_{\hat{\theta}} \sup_{p\in\mathcal{P}} \mathbb{E}_p[\Phi(\rho(\hat{\theta}(X_1, \cdots, X_n), \theta(p))]$, where the supremum is taken over distributions $p \in \mathcal{P}$ and the infimum over all estimators $\hat{\theta}$.

### 2.2 Local Differential Privacy and Private Minimax Risk

Since we will consider the sequential interactive and non-interactive local models in this paper, we follow the definitions in [3]. We assume that $\{Z_i\}_{i=1}^n$ are the private observations transformed from $\{X_i\}_{i=1}^n$ through some privacy mechanisms. We say that the mechanism is sequentially interactive, when it has the following conditional independence structure: $\{X_i, Z_1, \cdots, Z_{i-1}\} \mapsto Z_i, Z_i \perp\!\!\!\perp X_j \mid \{X_i, Z_1, \cdots, Z_{i-1}\}$ for all $j \neq i$ and $i \in [n]$, where $\perp\!\!\!\perp$ means independent relation. The full conditional distribution can be specified in terms of conditionals $Q_i(Z_i \mid X_i = x_i, Z_{1:i} = z_{i:i})$. The full privacy mechanism can be specified by a collection $Q = \{Q_i\}_{i=1}^n$.

When $Z_i$ is depending only on $X_i$, the mechanism is called non-interactive and in this case we have a simpler form for the conditional distributions $Q_i(Z_i \mid X_i = x_i)$. We now define local differential privacy by restricting the conditional distribution $Q_i$.

**Definition 1** ([3])**.** For a given privacy parameter $\epsilon > 0$, the random variable $Z_i$ is an $\epsilon$ sequentially locally differentially private view of $X_i$ if for all $z_1, z_2, \cdots, z_{i-1}$ and $x, x' \in \mathcal{X}$ we have the following for all the events $S$:

$$\frac{Q_i(Z_i \in S \mid X_i = x_i, Z_{1:i-1} = z_{1:i-1})}{Q_i(Z_i \in S \mid X_i = x_i', Z_{1:i-1} = z_{1:i-1})} \leq e^{\epsilon}.$$

We say that the random variable $Z_i$ is an $\epsilon$ non-interactively locally differentially private view of $X_i$ if

$$\frac{Q_i(Z_i \in S \mid X_i = x_i)}{Q_i(Z_i \in S \mid X_i = x_i')} \leq e^{\epsilon}.$$

We say that the privacy mechanism $Q = \{Q_i\}_{i=1}^n$ is $\epsilon$-sequentially (non-interactively) locally differentially private (LDP) if each $Z_i$ is a sequentially (non-interactively) locally differentially private view.

For a given privacy parameter $\epsilon > 0$, let $\mathcal{Q}_\epsilon$ be the set of conditional distributions that have the $\epsilon$-LDP property. For a given set of samples $\{X_i\}_{i=1}^n$, let $\{Z_i\}_{i=1}^n$ be the set of observations produced by any distribution $Q \in \mathcal{Q}_\epsilon$. Then, our estimator will be based on $\{Z_i\}_{i=1}^n$, that

is, $\hat{\theta}(Z_1, \cdots, Z_n)$. This yields a modified version of the minimax risk: $\mathcal{M}_n(\theta(\mathcal{P}), \Phi \circ \rho, Q) := \inf_{\hat{\theta}} \sup_{p \in \mathcal{P}} \mathbb{E}_p[\Phi(\rho(\hat{\theta}(Z_1, \cdots, Z_n), \theta(p))]$. This allows us to define functions that characterize the optimal rate of estimation in terms of privacy parameter $\epsilon$.

**Definition 2.** Given a family of distributions $\theta(\mathcal{P})$ and a privacy parameter $\epsilon > 0$, the $\epsilon$ sequential private minimax risk in the metric $\Phi \circ \rho$ is:

$$\mathcal{M}_n^{\text{Int}}(\theta(\mathcal{P}), \Phi \circ \rho, \epsilon) := \inf_{Q \in \mathcal{Q}_\epsilon} \mathcal{M}_n(\theta(\mathcal{P}), \Phi \circ \rho, Q),$$

where $\mathcal{Q}_\epsilon$ is the set of all $\epsilon$ sequentially locally differentially private mechanisms. Moreover, the $\epsilon$ non-interactive private minimax risk in the metric $\Phi \circ \rho$ is:

$$\mathcal{M}_n^{\text{Nint}}(\theta(\mathcal{P}), \Phi \circ \rho, \epsilon) := \inf_{Q \in \mathcal{Q}_\epsilon} \mathcal{M}_n(\theta(\mathcal{P}), \Phi \circ \rho, Q),$$

where $\mathcal{Q}_\epsilon$ is the set of all $\epsilon$ non-interactively locally differentially private mechanisms.

## 3 Limitations of Keeping Whole Dataset Private

We focus on the following distribution collection of samples $(x, y) \in \{+1, -1\}^p \times \mathbb{R}$:

$$\mathcal{P}_{1,p,C} = \{P_\theta \mid x \sim \text{Uniform}\{+1, -1\}^p, y = \langle \theta, x \rangle + \sigma, \\ \text{where} \sigma \text{ is the random noise satisfying the condition of } |\sigma| \leq C, \text{ and } \|\theta\|_2 \leq 1, \|\theta\|_0 \leq 1\}. \quad (1)$$

To show the limitations of the private minimax risk, we first give some intuition. Consider a raw data record $(x_i, y_i)$ which is sampled from some $p_\theta \in \mathcal{P}_{1,p,C}$. Suppose that we want to use a Gaussian or Laplacian mechanism on $(x_i, y_i)$ in order to make the algorithm locally differentially private. Then, due to sensitivity, the $\ell_1$ or $\ell_2$ norm of $(x_i, y_i)$ is a polynomial of $p$. The scale of the added random noise will also be a polynomial of $p$, which makes the final estimation error large.

The following theorem formally shows that for some fixed privacy parameter $\epsilon \in (0, 1)$, the $\epsilon$ non-interactive private minimax risk is lower bounded by a polynomial of the dimensionality $p$.

**Theorem 1.** For a given fixed privacy parameter $\epsilon \in (0, \frac{23}{35}]$, the $\epsilon$ non-interactive private minimax risk in the metric of $\|\cdot\|_2^2$ for the 1-sparse high dimensional sparse linear regression problem $\mathcal{P}_{1,p,2}$ needs to satisfy the following inequality, $\mathcal{M}_n^{\text{Nint}}(\theta(\mathcal{P}_{1,p,2}), \|\cdot\|_2^2, \epsilon) \geq \Omega(\min\{1, \frac{p \log p}{n\epsilon^2}\})$.

With the above theorem, our question now is to determine whether there are other factors in the local model that might allow us to avoid the polynomial dependency on $p$ in the estimation error.

We first consider the necessity of interaction in the model, since for some problems, such as convex Empirical Risk Minimization (ERM), there exists a large gap in the estimation error between the interactive and non-interactive local models [9]. The following theorem suggests that even if sequential interaction is allowed in the local model, the polynomial dependence on $p$ is still unavoidable. Note that sequential interaction is a commonly used model in LDP [3, 9].

**Theorem 2.** For a given fixed privacy parameter $\epsilon \in (0, \frac{23}{35}]$, the $\epsilon$ sequential private minimax risk in the metric of $\|\cdot\|_2^2$ for the 1-sparse high dimensional sparse linear regression problem $\mathcal{P}_{1,p,2}$ needs to satisfy the following inequality, $\mathcal{M}_n^{\text{int}}(\theta(\mathcal{P}_{1,p,2}), \|\cdot\|_2^2, \epsilon) \geq \Omega(\min\{1, \frac{p}{n\epsilon^2}\})$.

**Theorem 3.** Consider the loss function $L : \Theta \times \Theta \mapsto \mathbb{R}_+$, where $L(\theta, \theta') = |1^T(\theta - \theta')|$. Then, for any fixed $\epsilon \in (0, \frac{23}{35}]$, the $\epsilon$ sequential private minimax risk in the loss function $L$ for the 1-sparse high dimensional sparse linear regression problem $\mathcal{P}_{1,p,2}$ needs to satisfy the following inequality,

$$\mathcal{M}_n^{\text{int}}(\theta(\mathcal{P}_{1,p,2}), L, \epsilon) \geq \Omega(\min\{1, \sqrt{\frac{p}{n\epsilon^2}}\}). \quad (2)$$

## 4 Power of Keeping Responses Private

In this section, we consider the restricted case where only the responses or labels (*i.e.,* $\{y_i\}_{i=1}^n$) are required to be locally differentially private and all the observations $\{x_i\}_{i=1}^n$ are assumed to be

public. Preserving the privacy of the labels has been studied in [2, 1] for private PAC. We also note that keeping the responses private is related to some scenarios in physical sensor data and the sparse recovery problem, which have been studied in [6]. In this case, we can actually assume that $\{x_i\}_{i=1}^n \in (\{+1,-1\}^p)^n$ are fixed, and the collection of probability $\mathcal{P}_{s,p,C}$ in (1) is now reduced to the following model: $\mathcal{P}'_{s,p,C} = \{p_\theta(y_1,\cdots,y_n) \mid y_i = \langle\theta^*, x_i\rangle + \sigma_i$, where $\|\theta\|_0 \le s, \|\theta\|_2 \le 1$ and the random noise $|\sigma_i| \le C\}$.

The following theorem shows that, for every set of data $\{(x_i, y_i)\}_{i=1}^n$, if only $\{y_i\}_{i=1}^n$ need to be private, then there is an $(\epsilon, \delta)$ non-interactively locally differentially private algorithm DP-IHT, which yields a non-trivial upper bound on the squared $\ell_2$ norm of estimation error, see Algorithm 1. Before giving the theoretical analysis, we first show the assumption of the public dataset

**Algorithm 1** DP-Iterative Hard Thresholding

**Input**: Public dataset $\{x_i\}_{i=1}^n$, private $\{y_i\}_{i=1}^n \in P_{\theta^*}$, where $P_{\theta^*} \in \mathcal{P}'_{s^*,p,C}$, $\epsilon, \delta$ are privacy parameters, $T$ is the number of iteration, $\eta$ is the step size, and $s$ is the parameter to be specified. Set $\theta_0 = 0$.

**for** Each $i \in [n]$ **do**
  Denote $\tilde{y}_i = y_i + z_i$, where $z_i \sim \mathcal{N}(0, \sigma_1^2)$, $\sigma_1^2 = \frac{32C^2\ln(1.25/\delta)}{\epsilon^2}$.
**end for**
  **for** $t = 0, 1, \cdots, T-1$ **do**
    $\tilde{\theta}_{t+1} = \theta_t - \eta(\frac{1}{n}\sum_{i=1}^n(\tilde{y}_i - \langle x_i, \theta_t\rangle)x_i^T)$.
    $\theta'_{t+1} = \mathrm{Trunc}(\tilde{\theta}_{t+1}, s)$.
    $\theta_{t+1} = \arg_{\theta\in\mathbb{B}_1}\|\theta - \theta'_{t+1}\|_2^2$.
  **end for**
    Return $\theta_T$.

$X = (x_1^T, \cdots, x_n^T)^T \in \{+1,-1\}^{n\times p}$.

**Assumption 1.** $X$ satisfies the Restricted Isometry Property with parameter $2s + s^*$, where $s = 8s^*$. That is, for any $v \in \mathbb{R}^p$ with $\|v\|_0 \le 2s + s^*$, there exists a constant $\delta$ which satisfies $(1-\delta)\|v\|^2 \le \frac{1}{n}\|Xv\|_2^2 \le (1+\delta)\|v\|_2^2$.

Note that if $X = (x_1^T, \cdots, x_n^T)^T \sim \mathrm{Uniform}\{+1,-1\}^{n\times p}$, it satisfies the condition with probability ay least $1-\epsilon$ if $n \ge c\delta^{-2}(s^*\log p + \ln(1/\epsilon))$ with some universal constant $c$ (see Theorem 2.12 in [8]).

**Theorem 4.** For any $0 < \epsilon \le 1$ and $0 < \delta < 1$, Algorithm 1 is $(\epsilon, \delta)$ (non-interactively) locally differentially private for $\{y_i\}_{i=1}^n$. Moreover, if $\{y_i\}_{i=1}^n \in P_{\theta*}^n$, where $P_{\theta*} \in \mathcal{P}'_{s^*,p,C}$, and $X$ satisfies Assumption 1 with $0 < \delta \le \frac{2}{7}$, then by setting $s = 8s^*$ in Algorithm 1, there is an $\eta = \eta(\delta)$ which ensures that the output $\theta_T$ satisfies the following inequality

$$\|\theta_T - \theta^*\|_2 \le (\frac{1}{2})^T\|\theta^*\|_2 + O(\frac{C\log(1/\delta)\sqrt{s^*\log p}}{\sqrt{n}\epsilon}), \tag{3}$$

with probability at least $1 - \exp(-n) - \frac{2}{p}$.

From the above theorem, an immediate question is that whether the upper bound in Theorem 4 can be further improved. Unfortunately, the following theorem (adopted from [7]) indicates that the $\epsilon$ non-interactive local private minimax risk in the metric of $\|\cdot\|^2$ is lower bounded by $\Omega(\frac{C^2 s^*\log p}{n\epsilon^2})$, which means that the upper bound in Theorem 4 is tight.

**Theorem 5.** Under Assumption 1 and for a given fixed privacy parameter $\epsilon \in (0, \frac{23}{35}]$, the $\epsilon$ non-interactive local private minimax risk for the case of keeping $\{y_i\}_{i=1}^n$ locally private in the metric $\|\cdot\|^2$ satisfies the following inequality

$$\mathcal{M}_n^{\mathrm{Nint}}(\theta(\mathcal{P}'_{s,p,C}), \|\cdot\|_2^2, \epsilon) \ge \Omega\big(\min\{1, \frac{C^2 s\log\frac{p}{s}}{n\epsilon^2(1+\delta)}\}\big).$$

## References


[1] A. Beimel, K. Nissim, and U. Stemmer. Private learning and sanitization: Pure vs. approximate differential privacy. In *APPROX*, pages 363–378. Springer, 2013.

[2] K. Chaudhuri and D. Hsu. Sample complexity bounds for differentially private learning. In *COLT*, pages 155–186, 2011.

[3] J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Local privacy and statistical minimax rates. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 429–438. IEEE, 2013.

[4] J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521):182–201, 2018.

[5] J. C. Duchi and F. Ruan. The right complexity measure in locally private estimation: It is not the fisher information. *CoRR*, abs/1806.05756, 2018.

[6] A. McMillan and A. C. Gilbert. Local differential privacy for physical sensor data and sparse recovery. In *CISS*, pages 1–6. IEEE, 2018.

[7] G. Raskutti, M. J. Wainwright, and B. Yu. Minimax rates of estimation for high-dimensional linear regression over $\ell_q$-balls. *IEEE transactions on information theory*, 57(10):6976–6994, 2011.

[8] H. Rauhut. Compressive sensing and structured random matrices. *Theoretical foundations and numerical methods for sparse recovery*, 9:1–92, 2010.

[9] A. Smith, A. Thakurta, and J. Upadhyay. Is interaction necessary for distributed private learning? In *Security and Privacy (SP), 2017 IEEE Symposium on*, pages 58–77. IEEE, 2017.